Using initialized decadal hindcasts to assess simulations of natural variability and externally forced trends in the tropical Indo-Pacific Ocean

Amy Solomon

PSD/ESRL/NOAA and CIRES/University of Colorado

email:amy.solomon@noaa.gov

Revised paper submitted to Journal of Climate December 17, 2012

Abstract

This study assesses annual mean variability and 50-year trends in indices of tropical Indo-Pacific Ocean temperature and Pacific zonal wind stress in 4 CMIP5 ensembles of initialized decadal hindcasts to identify potential biases and systematic trends. We show that all models used in this analysis simulate statistically significant warming trends in Warm Pool and Cold Tongue near-surface temperatures. However, the Pacific near-equatorial zonal wind stress trends are inconsistent and not statistically significant at long lead-times. There is an indication that the zonal wind stress trends are proportional to the relative magnitude of the Warm Pool and Cold Tongue trends. All models show a warming trend in the Indian Ocean thermocline due to an unrealistic deepening of the thermocline. In the Pacific Ocean thermocline, all models have statistically significant trends but disagree as to whether it is a warming or cooling trend. Interestingly, we find no systematic relationship between near-equatorial zonal mean zonal wind stress trends and Pacific Ocean thermocline trends. PDFs are used to assess whether biases develop in annual mean distributions as a function of lead-time. In the Warm Pool and Cold Tongue regions, all ensembles show a bias towards a symmetric Gaussian distribution at long lead-times, which may cause an underestimate of predictability of a forced trend.

1. Introduction

The Intergovernmental Panel on Climate Change Fourth Assessment (AR4, Solomon et al. 2007) concluded that, as a result of an increase in well-mixed greenhouse gases all of North America is likely to warm in the 21st Century. In addition, annual mean precipitation is very likely to increase in Canada and the northeast USA, and likely to decrease in the southwest USA. To what extent are these climate variations due to the ocean's response to an increase in greenhouse gases? Studies such as Hoerling and Kumar (2003) attribute the persistent drought of 1998-2002 over the U.S. to sea surface temperatures (SSTs) in the Indian and tropical Pacific Oceans. These anomalous SSTs were in turn attributed to the ocean's response to an increase in greenhouse gases. These results are consistent with the atmospheric general circulation model (AGCM) studies of Barsugli et al. (2006) and Deser and Phillips (2009), among others.

This study focuses on trends in the Indo-Pacific Ocean, where natural multidecadal-to-centennial variations obscure our ability to identify the basin-wide response to an increase in greenhouse gases. For example, the sea surface temperatures (SST) in the Warm Pool region (18°N-18°S, 60°-165°E) have warmed by approximately 0.5°C since 1961 (Figure 1a). However, multidecadal trends in the Pacific and Indian Ocean thermocline have complicated the identification of how the tropical Indo-Pacific basin responds to an increase in greenhouse gases. Specifically, the tropical Indian Ocean exhibited a cooling upper thermocline trend and a warming surface trend from 1961-1998 (Trenary and Han 2008; Han et al. 2006; Lee 2004), which subsequently reversed after 1999 (Feng et al. 2010). These upper thermocline trends are primarily forced by variations in local Ekman pumping anomalies, which shoal or deepen the thermocline

(Alory et al. 2007), with changes in shallow overturning circulations playing a secondary role (Lee 2004; Schoenefeldt and Schott 2006). Similar multi-decadal trends have been observed in the western Pacific thermocline, where a pronounced cooling trend (along with a weakening of the shallow overturning circulations) was observed from 1960-1999 (McPhaden and Zhang 2002), which reversed thereafter (McPhaden and Zhang 2006; Lee and McPhaden 2008; Feng et al. 2010).

In the eastern equatorial Pacific, the identification of the response to an increase in greenhouse gases is even more challenging. For example, Solomon and Newman (2012) used a multivariate red noise model of El Nino/Southern Oscillation (ENSO) variability to demonstrate that centennial trends due to natural randomly occurring ENSO variability are larger than the observed trend in the eastern equatorial Pacific Ocean since 1900 in four SST reconstructions, indicating that if there is a trend due to external forcing in the eastern equatorial Pacific it is too small to be distinguished from natural variability. In addition, zonal mean zonal wind stress (τ_x) and sea level pressure (SLP) trends in the Pacific basin, which are dependent upon the relative strength of the trends in the Warm Pool and Cold Tongue (4°N-4°S, 170°E-70°W) regions (Meng et al. 2012), have reversed sign and strengthening since the 1990's (Merrifield 2011; Merrifield and Maltrud 2011). It has been suggested that the systematic changes in the Warm Pool SSTs may be driving these Indo-Pacific basin changes (Luo et al. 2012).

The World Climate Research Program's (WCRP) Working Group on Coupled Modeling has carried out a coordinated set of model experiments for the Intergovernmental Panel on Climate Change Fifth Assessment (AR5) that includes, for the first time, simulations of decadal climate prediction (hereafter referred to as

initialized decadal forecasts, see WCRP Joint Scientific Committee Session 29 Report). The ultimate goal of these simulations will be to provide policymakers with information on decadal time scales to assess possible consequences of climate change.

The Climate Model Intercomparison Project phases 3 and 5 (CMIP3 and CMIP5) organized an assessment of climate model simulations forced with an estimate of 20th-21st century external forcings (greenhouse gases, volcanic emissions, solar cycle variability, and aerosols) in order to verify and validate climate models used for projections of future climate change. These coupled climate model simulations produce a range of responses, in space and time, to anthropogenic radiative forcing (for example, see Solomon et al. (2011) Figure 8). Studies of CMIP3 ensembles clearly demonstrate that model uncertainty is the dominant source of uncertainty for projections of globally averaged surface temperature on decadal time scales (see Hawkins and Sutton 2009, 2011).

The extent to which initialization of the ocean state improves predictability on decadal time scales was first explored by Smith et al. (2007). This study found that initializing the Hadley Centre HadCM3 coupled climate model with an estimate of the observed ocean state produces a significant improvement in the forecast of the globally averaged mean surface temperature for a forecast period of 9 years. However, for forecasts over regions such as North America the initialized and uninitialized simulations had 9-year mean surface temperature anomaly root mean square errors (RMSEs) with similar patterns and equivalent magnitudes, highlighting the issue of limits on predictability due to model bias on these time scales. This study also found that the predictable climate change signal is determined entirely by external forcing after approximately 5 years, in agreement with studies such as Troccoli and Palmer (2007) but

in contrast to studies such as Keenlyside et al. (2008). These conclusions need to be tested with the CMIP5 decadal hindcasts.

2. Methods, Models and Data

In this study we use CMIP5 initialized decadal hindcasts to identify biases in natural and externally forced variability and to identify any potential systematic responses to an increase in greenhouse gases across ensembles from 4 coupled climate models. We focus on indices that have been used in past studies to define variability in the Indo-Pacific, specifically; the Warm Pool and Cold Tongue indices (previously defined), the Pacific τ_x index (5°N-5°S, 120°E-70°W), the Pacific Ocean thermocline index (125-250 meters depth, 5°N-5°S, 140°E-160°W), and the Indian Ocean thermocline index (75-275 meters depth, 9°S-16°S, 30°E-100°E). The areas used to form these indices are indicated in Figures 1 and 2.

We apply our analysis to four sets of initialized decadal hindcasts archived in the CMIP5 database (Taylor et al. 2012) that are initialized yearly from 1960-2009. Each start date has 10 ensemble members with perturbed initial conditions. These hindcasts take into account changes in external forcings such as greenhouse gases, solar activity, stratospheric aerosols associated with volcanic eruptions and anthropogenic aerosols. The first two ensembles use the UK Met Office coupled climate model HadCM3 configured with a horizontal resolution of 2.5^ox2.5^o in the atmosphere and 1.25^o in the ocean (Gordon et al. 2000). The HadCM3-i2 ensemble is anomaly initialized (observed anomalies and the models mean climate are used as initial conditions) and the HadCM3-i3 ensemble is initialized with full fields (observed anomalies and climate mean states are

used as initial conditions). The third ensemble uses the Canadian Centre for Climate Modelling and Analysis CanCM4 (Arora et al. 2011; Merryfield et al. 2011), and is full field initialized. The fourth ensemble uses the NOAA Geophysical Fluid Dynamics Laboratory CM2.1 (Delworth et al. 2006; Chang et al. 2012) and is anomaly initialized. All fields are interpolated to the HadCM3 2.5^{o}x2.5^{o} grid.

Annual mean anomalies are bias corrected as a function of lead-time, where the model forecast anomaly is calculated as $Y'_{j\tau} = Y_{j\tau} - \bar{Y}_{\tau}$, where $\bar{Y}_{\tau}$ is the ensemble-average forecast as a function of lead-time τ, $Y'_{j\tau}$ is the anomaly of the raw forecast with respect to the ensemble average, j is the starting year. $\bar{Y}_{\tau}$ is calculated as $\frac{1}{n}\sum_{j=1}^{n} Y_{j\tau}$. Figure 3 illustrates how forecasted trends are calculated as a function of lead-time in this study.

The verification datasets used in this study are three dimensional temperature fields from the European Centre for Medium-range Weather Forecasts (ECMWF) Ocean Reanalysis System 4 (ORAS4, Balmaseda et al. 2013) and the Geophysical Fluid Dynamics Laboratory Ensemble Coupled Data Assimilation V3.1 (ECDA, Chang et al. 2012). ORAS4 assimilates ECMWF EN3 v2a expendable bathythermograph (XBT) bias corrected temperature and salinity profiles, Archiving, Validation and Interpretation of Satellite Oceanographic data (AVISO) along track altimeter sea level anomalies and global trends, and ECMWF 40 Year Re-analysis (ERA-40), National Ocean and Atmospheric Administration (NOAA) Optimum Interpolation v2 and The National Centre for Ocean Forecasting Operational Sea Surface Temperature and Sea Ice Analysis SST and sea-ice using NEMOVAR in its 3D-var FGAT mode (Morgensen et al. 2012). Surface fluxes are from the ERA-40 atmospheric reanalysis, the ERA-Interim reanalysis,

and the operational ECMWF atmospheric analysis. ECDA applies an ensemble-based filtering algorithm to the GFDL coupled climate model CM2.1. ECDA assimilates World Ocean Database 2009 and global temperature–salinity profile program datasets temperature and salinity profiles, NOAA optimum interpolation SST v2 and HadISST SSTs, and NOAA National Centers for Environmental Prediction (NCEP) reanalysis winds, temperature and surface pressure. For zonal wind stress we use the ECDA alone since wind stress fields were not available for ORAS4.

Linear trends are calculated using the method of least squares linear regression. The Student's t-test (Wilks, 1995) is used to test the statistical significance of the trend. Trends are estimated to be significantly different from a zero trend when they exceed the 95% level. Both 95% and 98% significant trends are indicated in Tables 1-5.

Probability Distribution Functions (PDFs) are used in this study to assess potential biases in forecasts of natural variability in the Indo-Pacific basin. These biases in "climate noise" impact the predictability of robust linear trends. The one-tailed chi-square test is used to test whether PDFs come from different distributions. Chi-square is calculated with degrees of freedom equal to the number of bins minus two. Two distributions are estimated to come from different distributions when the p-values are less then 0.05 (there is less than a 5% chance that the two PDFs come from the same distribution). Hindcast PDFs are constructed using 500 annual means and detrended by removing the ensemble mean trend.

3. Results

a. Ensemble mean trends

Figures 1 and 2 show linear trends of the average of the ECDA and ORAS4 (referred to as AveObs from hereon) ocean temperature fields and the ECDA zonal wind stress fields from 1961-2009. The AveObs linear trend shows a warming trend throughout the tropical Indo-Pacific at 5m and a basin-wide cooling trend in both the Pacific and Indian Ocean equatorial thermoclines, with a basin-wide shoaling of the equatorial Pacific Ocean thermocline. The warming trend in the Cold Tongue region is less than the warming trend in the Warm Pool region (Tables 1 and 2), indicating a strengthening of the equatorial SST gradient. Trends averaged over the Cold Tongue and Pacific τ_x regions are not significant beyond the 95% level (Tables 2 and 3). The cooling trend in both the Pacific and Indian Ocean thermocline regions is statistically significant even though there has been a pronounced reversal in the trends since the end of the 1990's (Tables 4 and 5). Figure 2 shows a strengthening of the equatorial trades from 150°E-150°W and a weakening east of 150°W, indicating a divergent flow near 150°W, which is consistent with the shoaling of the Pacific Ocean thermocline near 150°W being due to reduced Ekman pumping. The cooling trend in the Indian Ocean thermocline peaks between 16-8°S but weaker cooling trends are seen to extend from to 18°S to 25°N.

In Figure 4 we show the Warm Pool indices for lead-times of 1, 2, 6, and 10 years. The close correspondence between the observed warming trend and the forecasted trend is seen for lead-times up to 10 years. However, as can be seen clearly in Table 1, there is a systematic increase in the warming trend for longer lead-times for all 4 ensembles. All ensemble mean trends in the Warm Pool region are statistically significant beyond the 98% level.

As stated previously, the observed trend in the Cold Tongue region is not statistically significant. This is the case for the forecasted trend as well as for lead-times less than 3 years. However, as seen in Table 2 and Figure 5, all ensembles produce a warming trend that is significant beyond the 98% level for lead-times greater than 2 years. These warming trends are 45-95% larger than the observed Cold Tongue trend. These warming trends may be due to greater statistical significance due to less aliasing from the observed anomalies as the model becomes independent of the initial conditions or due to the model developing unrealistic trends for longer lead-times, given the large uncertainty in the observed Cold Tongue region it is not possible to determine which is occurring. It is interesting to note that the HadCM3 i2 and i3 ensembles produce very similar trends even though one is anomaly initialized and the other is full field initialized. Considering the large bias correction needed for the HadCM3 i3 ensemble (see Kim et al. 2012), this is an indication of the linearity of the response to external forcing in this model.

Even though both the Warm Pool and Cold Tongue ensemble mean trends are significant for all ensembles for lead-times greater than 2 years, there is no consistency in the modeled Pacific τ_x trends (Table 3 and Figure 6). For example, Table 3 shows that only CanCM4 has a trend averaged over lead-times of 6-10 years that is significant beyond the 95% level. This trend is a weakening of the zonal mean equatorial trades, while the other three models produce a statistically insignificant strengthening of the equatorial trades. However, even though the majority of these trends are not statistically significant, there is an indication that the Pacific τ_x trends are related to the relative

magnitude of the Warm Pool and Cold Tongue trends (τ_x trends $\propto$ Cold Tongue-Warm Pool trends).

All ensembles produce statistically significant trends in the Pacific Ocean thermocline region for long lead-times (Table 4 and Figure 7). However there is no agreement in the sign of the forecasted trend. The HadCM3 models produce a warming trend while the CanCM4 and GFDL models produce a cooling trend that is 0.22-0.43 times smaller than the observed trend. There is no systematic relationship between the forecasted Pacific τ_x trends and the Pacific Ocean thermocline trends. Specifically, there is no consistency between the CanCM4 and GFDL Pacific τ_x trends even though at long lead-times both show a basinwide cooling trend in the Pacific Ocean thermocline similar to the observed trend in Figure 1 (results not shown).

In the Indian Ocean thermocline region, all ensembles show a cooling trend for lead-times less than 3 years, similar to observations (Table 5 and Figure 8). However, the HadCM3 i3 trends for lead-times less than 3 years are not statistically significant. For longer lead-times all ensembles develop warming trends, all of which are statistically significant beyond the 98% level except for the GFDL ensemble mean trend. These warming trends are due to an unrealistic deepening of the Indian Ocean thermocline across the basin that increases with lead-time (seen in Table 5). For example, for averages over 6-10 year lead-times, HadCM3 i2 warming trends extend to 300 meters, HadCM3 i3 trends extend to 300 meters, CanCM4 warming trends extend to 200 meters, and GFDL warming trends extend to 250 meters (results not shown), while observed trends in the equatorial east Indian Ocean do not extend below 150 meters (Figure 1).

b. Annual mean variability

Figure 9 shows the distribution of detrended annual mean observations in the Warm Pool region compared to forecasts at lead-times of 2, 6, and 10 years. All ensembles at a lead-time of 2 years have distributions that are statistically indistinguishable from the observations. However, at a lead-time of 10 years all ensembles except GFDL have PDFs that come from different distributions than the observations. Specifically, all ensembles show a distribution that is more symmetric than the observations, which show more frequent extreme cold years than warm years. This may be a sampling issue with the limited number of observations or there may be an asymmetry in the observed distribution that is not captured in the models (also seen in the Pacific Ocean thermocline PDFs, discussed below). The similarity between the year 2 lead-time distribution and observations indicate that it is the later since this PDF was constructed with 500 samples. In any case, the PDFs at a lead-time of 6 and 10 years show a bias towards a symmetric Gaussian distribution (Gaussian not shown).

Observations in the Cold Tongue region show an asymmetric distribution, with a more frequent occurrence of extreme warm years than cold years (Figure 10), due to the large amplitude of El Niño events relative to La Niña events. All ensembles at a lead-time of 2 years have a distribution that is indistinguishable from the observed distribution. The HadCM3 ensembles at a lead-time of 10 years have a distribution that is indistinguishable from the observed distribution. The CanCM4 and GFDL distributions become more Gaussian with lead-time; at a lead-time of 10 years these two ensembles produce a similar number of extreme warm and cold years. However, this tendency to a more symmetric distribution with longer lead-times is seen in all four models indicating that

the statistics of ENSO events evolve from an asymmetric distribution to a distribution where La Niña and El Niño events have similar amplitudes.

For the Pacific τ_x distributions, none of the forecasted distributions at a lead-time of 2 years come from the same distribution as observations (Figure 11). However, the CanCM4 and the GFDL distributions at a lead-time of 10 years are indistinguishable from the observed distribution. These two models simulate the observed skew towards strengthened equatorial trades while the HadCM3 ensembles have distributions that are independent of lead-time and underestimate the frequency of both extreme positive and negative years, potentially overestimated the predictability of a forced trend.

In the Pacific Ocean thermocline region, all distributions from the hindcasts come from different distributions than the observations except for the GFDL distribution for a lead-time of 10 years, primarily due to the overestimate of cold years and a tendency towards a more symmetric distribution (Figure 12). Interestingly, the GFDL forecast at a lead of 10 years captures the peak skewed towards positive values and the long tail for extreme cold years quite well. Again the underestimate of the frequency of extreme cold years in the other three ensembles may overestimate the predictability of a forced trend.

In the Indian Ocean thermocline region, only HadCM3 i2 and CanCM4 distributions at a lead-time of 2 years are indistinguishable from the observed distribution (Figure 13). The GFDL distribution at a lead-time of 10 years shows the same skew towards positive values and the long tail due to the frequency of extreme cold years as the Pacific Ocean thermocline index that is not seen in the observed distribution. All distributions at a lead-time of 10 years show a broader distribution than observations,

potentially underestimating the predictability of a forced trend due to the overestimate of both extreme warm and cold years, consistent with the results of Corti et al. (2012).

4. Summary and Conclusions

In summary, the main findings of this study are:

1) All ensembles produce a statistically significant warming trend in the Warm Pool region that systematically increases as a function of lead-time. All ensembles show a bias towards a symmetric Gaussian distribution in the Warm Pool region, whereas the observed distribution shows more frequent extreme cold years than warm years.
2) All ensembles produce a warming trend in the Cold Tongue region that is significant beyond the 98% level for lead-times greater than 2 years. All ensembles have a tendency to develop more symmetric distribution with longer lead-times indicating that the statistics of ENSO events evolve from an asymmetric distribution to a distribution where La Niña and El Niño events have similar amplitudes.
3) There is an indication that the Pacific τ_x trends are related to the relative magnitude of the Warm Pool and Cold Tongue trends.
4) There is no systematic relationship between the forecasted Pacific τ_x trends and the Pacific Ocean thermocline trends.
5) In the Pacific Ocean thermocline, statistically significant trends are only found across ensembles for 6-10 year lead-time averages, however two of the ensembles produce a warming trend and two of the ensembles produce a cooling trend.
6) Warming trends develop in the Indian Ocean thermocline region at long lead-times due to an unrealistic deepening of the Indian Ocean thermocline across the basin. The

GFDL distribution at a lead-time of 10 years shows the same skew towards positive values and the long tail due to the frequency of extreme cold years in both the Indian and Pacific Ocean thermocline regions. However, all distributions at a lead-time of 10 years show a broader distribution than observations, potentially underestimating the predictability of a forced trend due to the overestimate of both extreme warm and cold years.

The different relationships between near-surface and subsurface temperature trends in each of the ensembles is an indication that models may reproduce observed surface trends but produce unrealistic subsurface trends due to an unrealistic heat balance in the upper ocean. For example, all ensembles produce systematic trends in the Warm Pool region, similar to observations, but fail to reproduce the observed shoaling of the Indian Ocean equatorial thermocline. This is an indication that the ensembles overestimate the "climate change commitment" due to warming trends in the upper ocean.

In addition, the Pacific Ocean thermocline trends are much more uncertain than the near-surface trends. For example, the GFDL ensemble does not have statistically significant trends in the Pacific Ocean thermocline for lead times up to 10 years (however, an average over 6-10 year lead times is significant), while both Cold Tongue and Warm Pool trends are significant beyond the 98% level after a lead time of 6 years. This result highlights the difficulty of identifying externally forced trends in the short record of observed subsurface fields, as demonstrated in the reversal of trends in the Pacific and Indian Oceans since the late 1990's.

This study highlights the large uncertainties in zonal mean zonal wind stress trends. For example, all ensembles have trends in the Cold Tongue, Warm Pool, and Pacific thermocline regions that are significant beyond the 98% level but Pacific τ_x trends that are only significant at the 95% level for one ensemble when averaged over lead-times of 6-10 years. For the ensemble that does produce a statistically significant Pacific τ_x trend for an average over lead-times of 6-10 years (CanCM4), there is a consistent relationship between a weakening of the trades and a shoaling of the basin-wide equatorial Pacific thermocline. There is also an indication that the sign of the Pacific τ_x trends are proportional to the relative magnitude of trends in the Warm Pool and Cold Tongue regions. But given the large variability in the zonal wind stress, larger ensembles (and additional models) are needed to confirm these speculations.

A detailed heat budget analysis of the upper ocean needs to be done to determine the relative importance of local and remote forcing in the Indo-Pacific basin-wide response to external forcing in these models. A heat budget analysis will be useful in identifying the relative roles of specific processes, such as the ocean thermostat in the eastern equatorial Pacific, in each model's response to external forcing. If more fields from the decadal hindcasts become available to the scientific community this heat budget analysis will be done in a future study.

Acknowledgements: The author thanks M. Newman, C. Deser, and three anonymous reviewers for comments on this study. We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modeling groups (listed in this paper) for producing and making available their model output. For CMIP the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. This work was supported by grants from the NOAA OAR CVP program and NSF AGS 1125561.

References

Alory, G., S. Wijffels, and G. Meyers, 2007: Observed temperature trends in the Indian Ocean over 1960–1999 and associated mechanisms, *Geophys. Res. Lett.*, **34**, L02606, doi:10.1029/2006GL028044.

Arora, V., and Coauthors, 2011: Carbon emission limits required to satisfy future representative concentration pathways of greenhouse gases. *Geophys. Res. Lett.*, **38**, L05805, doi:10.1029/2010GL046270.

Balmaseda, M. A., K. Morgensen, and A. Weaver, 2013: Evaluation of the ECMWF Ocean Reanalysis ORAS4. *Q. J. Roy. Met*. In press.

Barsugli, J. J., S.-I. Shin, and P. D. Sardeshmukh, 2006: Sensitivity of global warming to the pattern of tropical ocean warming. *Clim. Dyn.*, **27**, 483-492.

Chang, Y.-S., S. Zhang, A. Rosati, T. Delworth, and W. F. Stern, 2012: An assessment of oceanic variability for 1960-2010 from the GFDL ensemble coupled data assimilation, *Clim. Dyn*.. In press.

Corti, S., A. Weisheimer, T. N. Palmer, F. J. Doblas-Reyes, and L. Magnusson, 2012: Reliability of decadal predictions, *Geophys. Res. Lett.*, **39**, L21712, doi:10.1029/2012GL053354.

Delworth, T. L., and Coauthors, 2006: GFDL's CM2 global coupled climate models. part I: Formulation and simulation characteristics. *J. Clim*., **19**, 643–674.

Deser, C., and A. S. Phillips, 2009: Atmospheric circulation trends, 1950-2000: The relative roles of sea surface temperature forcing and direct atmospheric radiative forcing. *J. Clim.*, **22**, 396-413.

Doblas-Reyes, F. J., M. A. Balmaseda, A. Weisheimer, and T. N. Palmer, 2011: Decadal climate prediction with the European Centre for Medium-Range Weather Forecasts coupled forecast system: Impact of ocean observations, *J. Geophys. Res.*, **116**, D19

Feng, M., M. J. McPhaden, and T. Lee, 2010: Decadal variability of the Pacific subtropical cells and their influence on the southeast Indian Ocean. *Geophys. Res. Lett.* **37**, L09606.

Gordon, C., C. Cooper, C. A. Senior, H. Banks, J. M. Gregory, T. C. Johns, J. F. B. Mitchell, and R. A. Wood, 2000: The simulation of SST, sea ice extents and ocean heat transport in a version of the Hadley Centre coupled model without flux adjustments. *Clim. Dyn*. **16**, 147–168.

Han, W., G. A. Meehl, and A. Hu, 2006: Interpretation of tropical thermocline cooling in the Indian and Pacific Oceans during recent decades. *Geophys. Res. Lett*., **33**, L23615, doi:10.1029/2006GL027982.

Hawkins, E., and R. T. Sutton, 2009: The potential to narrow uncertainty in regional climate predictions. *Bull. Am. Meteorol.. Soc.*, **90**, 1095-1107. ISSN 1520-0477 doi: 10.1175/2009BAMS2607.1.

Hawkins, E., and R. T. Sutton, 2011: The potential to narrow uncertainty in projections of regional precipitation change. *Clim. Dyn.*, **37**, 407-418. ISSN 1432-0894 doi: 10.1007/s00382-010-0810-6.

Hoerling, M., and A. Kumar, 2003: The Perfect Ocean for Drought. *Science*, **299**, 691-694.

Keenlyside, N. S., M. Latif, J. Jungclaus, L. Kornblueh, and E. Roeckner, 2008: Advancing decadal-scale climate prediction in the North Atlantic sector. *Nature*, **453**, 84-88.

Kim, H. M., P. J. Webster, and J. A. Curry, 2012: Evaluation of short-term climate change prediction in multi-model CMIP5 decadal hindcasts, *Geophys. Res. Lett.*, **39**, L10701, doi:10.1029/2012GL051644

Lee, T., 2004: Decadal weakening of the shallow overturning circulation in the South Indian Ocean. *Geophys. Res. Lett.*, **31**, L18305, doi:10.1029/2004GL020884.

Lee, T., and M. J. McPhaden, 2008: Decadal phase change in large-scale sea level and winds in the Indo-Pacific region at the end of the 20th century. *Geophys. Res. Lett.*, **35**, L01605, doi:10.1029/2007GL032419.

McPhaden, M. J., and D. Zhang, 2002: Slowdown of the meridional overturning circulation in the upper Pacific Ocean. *Nature*, **415**, doi: 10.1038/415603a, 603–608.

McPhaden, M. J., and D. Zhang, 2004: Pacific Ocean circulation rebounds. *Geophys. Res. Lett.*, 31, L18301, doi: 10.1029/2004GL020727.

Merrifield, M. A., 2011: A shift in western tropical Pacific sea-level trends during the 1990s. *J. Clim.,* 24, 4126–4138.

Merrifield, M. A., and M. E. Maltrud, 2011: Regional sea level trends due to a Pacific trade wind intensification. *Geophys. Res. Lett.,* 38, L21605.

Merryfield, W. J., W. Lee, G. J. Boer, V. V. Kharin, J. F. Scinocca, G. F. Flato, R. S. Ajayamohan, Y. Tang, and S. Polavarapu, 2012: The Canadian Seasonal to Interannual 1 Prediction System. Part I: Models and initialization*, Mon. Wea. Rev.*, in press.

Morgensen, K., M. A. Balmaseda, and A. Weaver, 2012: The NEMOVAR ocean data assimilation system as implemented in the ECMWF ocean analysis for System 4. ECMWF Technical Memorandum 668. 59 pages.

Schoenefeldt, R., and F. A. Schott, 2006: Decadal variability of the Indian Ocean cross-equatorial exchange in SODA, *Geophys. Res. Lett.*, **33**, L08602, doi:10.1029/2006GL025891.

Smith, D. M., S. Cusack, A. W. Colman, C. K. Folland, G. R. Harris, and J. M. Murphy, 2007: Improved surface temperature prediction for the coming decade from a global climate model. *Science*, **317**, 796-799.

Solomon, A., and M. Newman, 2012: Reconciling disparate 20th century Indo-Pacific ocean temperature trends in the instrumental record. *Nature Climate Change*, **2**, 691-699, doi:10.1038/NCLIMATE1591.

Solomon, A., and the US CLIVAR Working Group on Decadal Predictability, 2011: Distinguishing the roles of natural and anthropogenically forced decadal climate variability: Implications for prediction, *Bull. Am. Meteorol.. Soc*., **92**, 141-156, doi:10.1175/2010BAMS2962.1.

Solomon, S., et al., Eds., Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change (Cambridge Univ. Press, Cambridge, 2007).

Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An Overview of CMIP5 and the Experiment Design. *Bull. Amer. Meteor. Soc.*, **93**, 485–498. doi: http://dx.doi.org/10.1175/BAMS-D-11-00094.1

Trenary, L., and W. Han, 2008: Causes of decadal subsurface cooling in the tropical Indian Ocean during 1961-2000. *Geophys. Res. Lett*., **35**, L17602, doi:10.1029/2008GL034687.

Troccoli, A., and T. N. Palmer, 2007: Ensemble decadal predictions from analysed initial conditions. *Phil. Trans. R. Soc. A*., **365**, 2179-2191.

Wilks, D. S. 1995: *Statistical Methods in the Atmospheric Sciences*. Academic Press, New York, p. 145-157.

Zhang, D., and M. J. McPhaden, 2006: Decadal variability of the shallow Pacific meridional overturning circulation: relation to tropical sea surface temperatures in observations and climate change models. *Ocean Modelling,* **15***,* 250-273.

Table Captions:

Table 1: Warm Pool index ensemble mean 50-year linear trends as a function of lead-time, in units of °C/year. Bold font indicates that the trends are significant beyond the 95% level. Underlined numbers indicate that the trends are significant beyond the 98% level. In Tables 1-5 orange shading indicates systematic increases or decreases in trends as a function of lead-time. Green shading indicates large differences in trends across the ensembles for the average over year 6-10 lead-times. Blue shading indicates systematic over or under estimates across ensembles for the average over year 6-10 lead-times.

Table 2: Cold Tongue index ensemble mean 50-year linear trends as a function of lead-time, in units of °C/year. Bold font indicates that the trends are significant beyond the 95% level. Underlined numbers indicate that the trends are significant beyond the 98% level.

Table 3: Pacific τ_x index ensemble mean 50-year linear trends as a function of lead-time, in units of dPa/year. Bold font indicates that the trends are significant beyond the 95% level. Underlined numbers indicate that the trends are significant beyond the 98% level.

Table 4: Pacific Ocean thermocline index ensemble mean 50-year linear trends as a function of lead-time, in units of °C/year. Bold font indicates that the trends are significant beyond the 95% level. Underlined numbers indicate that the trends are significant beyond the 98% level.

Table 5: Indian Ocean thermocline index ensemble mean 50-year linear trends as a function of lead-time, in units of °C/year. Bold font indicates that the trends are

significant beyond the 95% level. Underlined numbers indicate that the trends are significant beyond the 98% level.

Figure Captions:

Figure 1: 1961-2009 tropical Indo-Pacific upper ocean trends calculated from the ECDA and ORAS4. A) At 5 m. Boxes indicate regions used to form Warm Pool (18°N-18°S, 60°-165°E) and Cold Tongue (4°N-4°S, 170°E-70°W) indices. B) Zonally averaged across the Indian Ocean. Box indicates region used to form the Indian Ocean thermocline index (75-275 meters depth, 9°S-16°S, 30°E-100°E). C) Along the Equator. Box indicates the region used to form the Pacific Ocean thermocline index (125-250 meters depth, 5°N-5°S, 140°E-160°W). Contour interval equal to 0.004°C/year.

Figure 2: 1961-2009 ECDA τ_x linear trend, in units of Pa/year. Box indicates the region used to form the τ_x index (5°N-5°S, 120°E-70°W).

Figure 3: Schematic of forecasted trends in initialized 10-year hindcasts. Hindcasts are initialized the year preceding forecast year-1 on November 1. To illustrate how the forecast years are defined, forecast year-1 is highlighted in red and forecast year-6 is highlighted in blue.

Figure 4: HadCM3 i2 (left) and HadCM3 i3 (right) Warm Pool index, in units of °C. Ensemble mean (thick black line) and ensemble spread (gray shading) for lead-time of 1, 2, 6, and 10 years. Observations shown with red line. Numbers show linear trend of the ensemble mean.

Figure 4 continued: CanCM4 (left) and GFDL (right) Warm Pool index.

Figure 5: HadCM3 i2 (left) and HadCM3 i3 (right) Cold Tongue index, in units of °C. Ensemble mean (thick black line) and ensemble spread (gray shading) for lead-time of 1,

2, 6, and 10 years. Observations shown with red line. Numbers show linear trend of the ensemble mean.

Figure 5 continued: CanCM4 (left) and GFDL (right) Cold Tongue index.

Figure 6: HadCM3 i2 and HadCM3 i3 Pacific τ_x index, in units of dPa. Ensemble mean (thick black line) and ensemble spread (gray shading) for lead-time of 1, 2, 6, and 10 years. Observations shown with red line. Numbers show linear trend of the ensemble mean.

Figure 6 continued: CanCM4 and GFDL Pacific τ_x index.

Figure 7: HadCM3 i2 (left) and HadCM3 i3 (right) Pacific Ocean thermocline index, in units of °C. Ensemble mean (thick black line) and ensemble spread (gray shading) for lead-time of 1, 2, 6, and 10 years. Observations shown with red line. Numbers show linear trend of the ensemble mean.

Figure 7 continued: CanCM4 (left) and GFDL (right) Pacific Ocean thermocline index.

Figure 8: HadCM3 i2 (left) and HadCM3 i3 (right) Indian Ocean thermocline index, in units of °C. Ensemble mean (thick black line) and ensemble spread (gray shading) for lead-time of 1, 2, 6, and 10 years. Observations shown with red line. Numbers show linear trend of the ensemble mean.

Figure 8 continued: CanCM4 (left) and GFDL (right) Indian Ocean thermocline index.

Figure 9: PDFs of detrended annual mean Warm Pool indices for lead-times of 2 years (green), 6 years (red), and 10 years (blue), in units of °C. PDFs of detrended observed

annual means shown with black dashed line. A) HadCM3 i2. B) HadCM3 i3. C) CanCM4. D) GFDL.

Figure 10: PDFs of detrended annual mean Cold Tongue indices for lead-times of 2 years (green), 6 years (red), and 10 years (blue), in units of °C. PDFs of detrended observed annual means shown with black dashed line. A) HadCM3 i2. B) HadCM3 i3. C) CanCM4. D) GFDL.

Figure 11: PDFs of detrended annual mean Pacific τ_x indices for lead-times of 2 years (green), 6 years (red), and 10 years (blue), in units of dPa. PDFs of detrended observed annual means shown with black dashed line. A) HadCM3 i2. B) HadCM3 i3. C) CanCM4. D) GFDL.

Figure 12: PDFs of detrended annual mean Pacific Ocean thermocline indices for lead-times of 2 years (green), 6 years (red), and 10 years (blue), in units of °C. PDFs of detrended observed annual means shown with black dashed line. A) HadCM3 i2. B) HadCM3 i3. C) CanCM4. D) GFDL.

Figure 13: PDFs of detrended annual mean Indian Ocean thermocline indices for lead-times of 2 years (green), 6 years (red), and 10 years (blue), in units of °C. PDFs of detrended observed annual means shown with black dashed line. A) HadCM3 i2. B) HadCM3 i3. C) CanCM4. D) GFDL.

Warm Pool Trends	AveObs	HadCM3 I2	HadCM3 I3	CanCM4	GFDL CM2.1
Forecast Year 1	**1.06e-2**	**1.09e-2**	**1.15e-2**	**0.71e-2**	**1.43e-2**
Forecast Year 2		**1.17e-2**	**1.16e-2**	**0.92e-2**	**1.60e-2**
Forecast Year 6		**1.26e-2**	**1.21e-2**	**1.53e-2**	**1.81e-2**
Forecast Year 10		**1.44e-2**	**1.39e-2**	**1.61e-2**	**1.86e-2**
Forecast Years 6-10		**1.34e-2**	**1.34e-2**	**1.56e-2**	**1.84e-2**

Table 1: Warm Pool index ensemble mean 50-year linear trends as a function of lead-time, in units of °C/year. Bold font indicates that the trends are significant beyond the 95% level. Underlined numbers indicate that the trends are significant beyond the 98% level. In Tables 1-5 orange shading indicates systematic increases or decreases in trends as a function of lead-time. Green shading indicates large differences in trends across the ensembles for the average over year 6-10 lead-times. Blue shading indicates systematic over or under estimates across ensembles for the average over year 6-10 lead-times.

Cold Tongue Trends	AveObs	HadCM3 I2	HadCM3 I3	CanCM4	GFDL CM2.1
Forecast Year 1	8.73e-3	8.04e-3	5.28e-3	-6.92e-3	-1.01e-3
Forecast Year 2		5.85e-3	5.01e-3	**10.07e-3**	-0.55e-3
Forecast Year 6		**8.06e-3**	**8.17e-3**	**17.59e-3**	**16.98e-3**
Forecast Year 10		**12.70e-3**	**12.91e-3**	**16.55e-3**	**16.95e-3**
Forecast Years 6-10		**10.00e-3**	**12.85e-3**	**16.62e-3**	**17.73e-3**

Table 2: Cold Tongue index ensemble mean 50-year linear trends as a function of lead-time, in units of °C/year. Bold font indicates that the trends are significant beyond the 95% level. Underlined numbers indicate that the trends are significant beyond the 98% level.

Pacific τ_x Trends	ECDA	HadCM3 I2	HadCM3 I3	CanCM4	GFDL CM2.1
Forecast Year 1	-2.46e-4	-1.98e-4	-2.70e-4	**-9.30e-4**	-3.16e-4
Forecast Year 2		-3.40e-4	-3.29e-4	-2.60e-4	**-13.80e-4**
Forecast Year 6		**-3.37e-4**	**-2.87e-4**	2.85e-4	-1.42e-4
Forecast Year 10		-1.23e-4	-0.19e-4	0.79e-4	-2.70e-4
Forecast Years 6-10		-2.77e-4	-1.31e-4	**1.35e-4**	-1.56e-4

Table 3: Pacific τ_x index ensemble mean 50-year linear trends as a function of lead-time, in units of dPa/year. Bold font indicates that the trends are significant beyond the 95% level. Underlined numbers indicate that the trends are significant beyond the 98% level.

Pacific Ocean Thermocline Trends	AveObs	HadCM3 I2	HadCM3 I3	CanCM4	GFDL CM2.1
Forecast Year 1	**-2.27e-2**	-0.71e-2	-0.93e-2	0.17e-2	-0.81e-2
Forecast Year 2		0.25e-2	0.30e-2	-0.33e-2	0.78e-2
Forecast Year 6		**1.14e-2**	**0.99e-2**	**-1.18e-2**	-0.29e-2
Forecast Year 10		**1.28e-2**	**0.76e-2**	**-0.76e-2**	-0.02e-2
Forecast Years 6-10		**1.43e-2**	**0.86e-2**	**-0.98e-2**	**-0.49e-2**

Table 4: Pacific Ocean thermocline index ensemble mean 50-year linear trends as a function of lead-time, in units of °C/year. Bold font indicates that the trends are significant beyond the 95% level. Underlined numbers indicate that the trends are significant beyond the 98% level.

Indian Ocean Thermocline Trends	AveObs	HadCM3 I2	HadCM3 I3	CanCM4	GFDL CM2.1
Forecast Year 1	**-1.05e-2**		-0.38e-2	**-1.22e-2**	**-1.17e-2**
Forecast Year 2			-0.56e-2	**-0.82e-2**	**-1.60e-2**
Forecast Year 6		**0.57e-2**	0.28e-2	0.18e-2	-0.01e-2
Forecast Year 10		**1.08e-2**	**0.95e-2**	**0.81e-2**	0.51e-2
Forecast Years 6-10		**0.76e-2**	**0.69e-2**	**0.48e-2**	0.26e-2

Table 5: Indian Ocean thermocline index ensemble mean 50-year linear trends as a function of lead-time, in units of °C/year. Bold font indicates that the trends are significant beyond the 95% level. Underlined numbers indicate that the trends are significant beyond the 98% level.

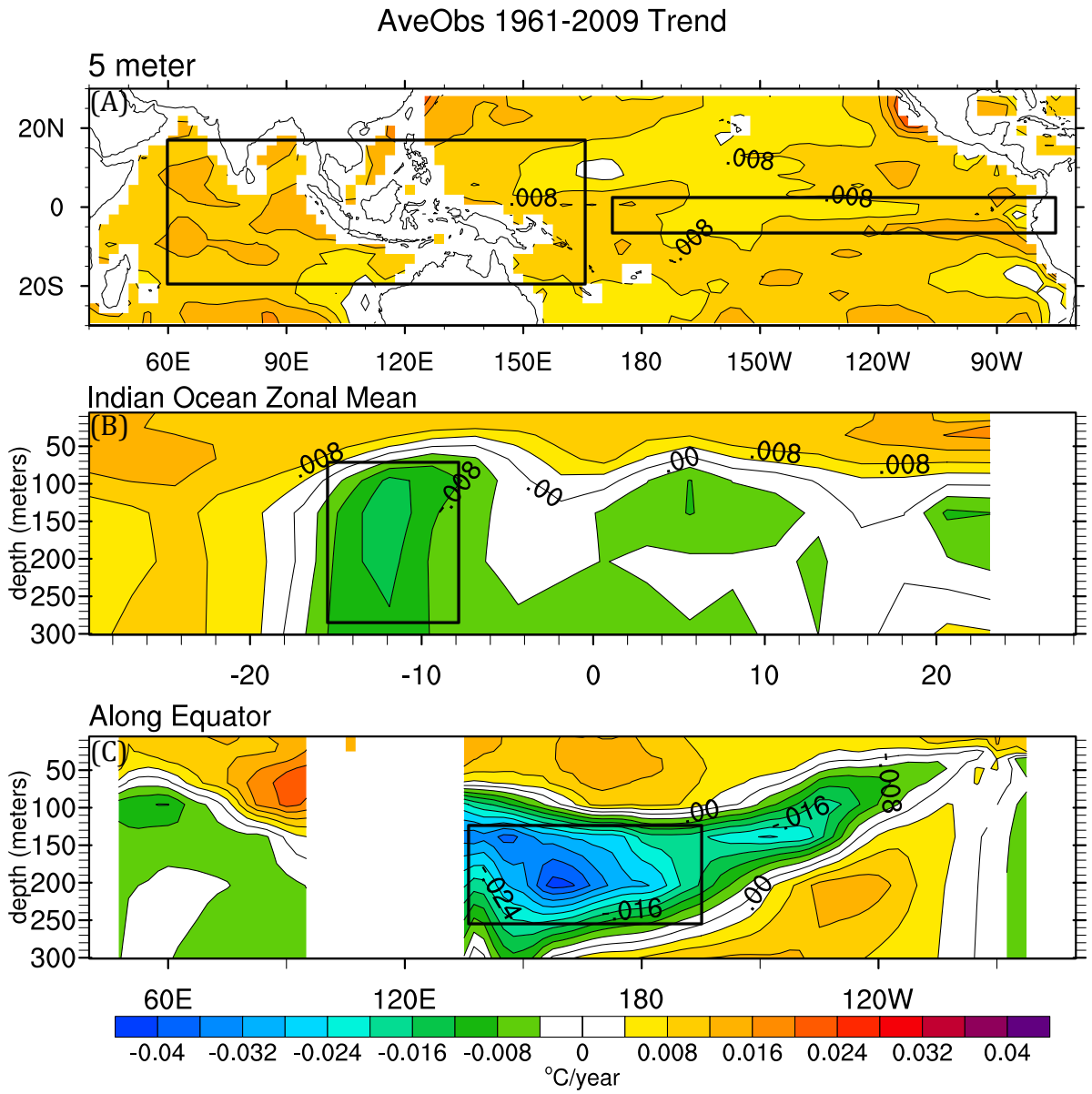

Figure 1: 1961-2009 tropical Indo-Pacific upper ocean trends calculated from the ECDA and ORAS4. A) At 5 m. Boxes indicate regions used to form Warm Pool (18°N-18°S, 60°-165°E) and Cold Tongue (4°N-4°S, 170°E-70°W) indices. B) Zonally averaged across the Indian Ocean. Box indicates region used to form the Indian Ocean thermocline index (75-275 meters depth, 9°S-16°S, 30°E-100°E). C) Along the Equator. Box indicates the region used to form the Pacific Ocean thermocline index (125-250 meters depth, 5°N-5°S, 140°E-160°W). Contour interval equal to 0.004°C/year.

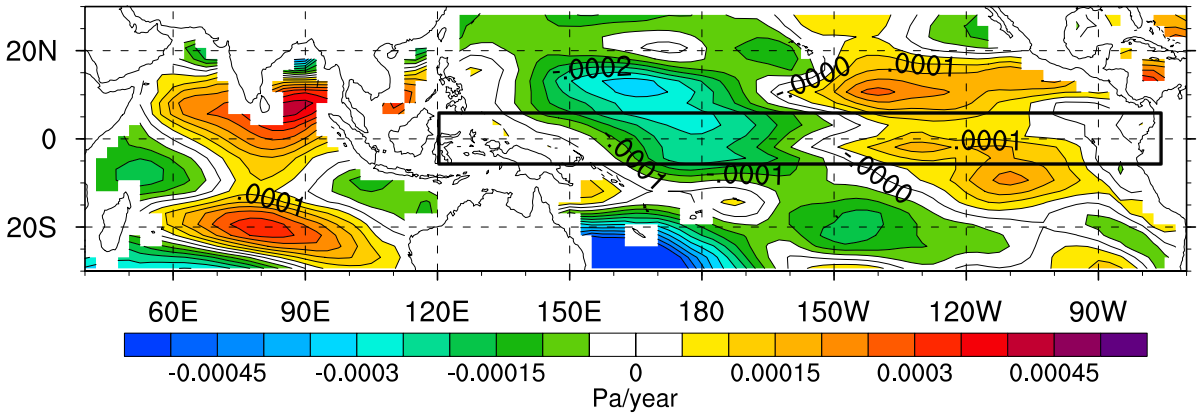

Figure 2: 1961-2009 GFDL ECDA τ_x linear trend, in units of Pa/year. Box indicates the region used to form the τ_x index (5°N-5°S, 120°E-70°W).

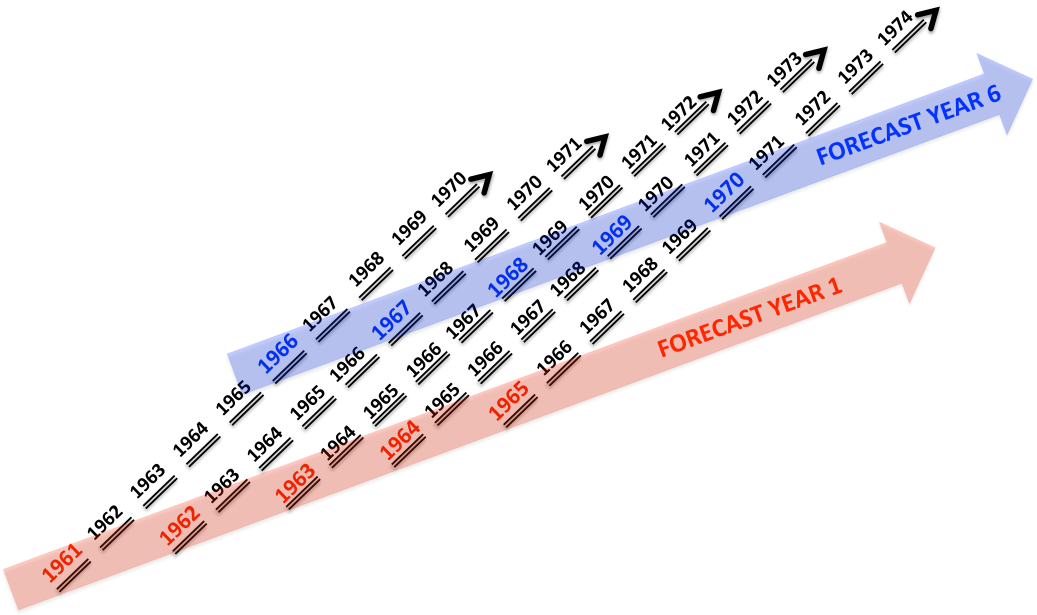

Figure 3: Schematic of forecasted trends in initialized 10-year hindcasts. Hindcasts are initialized the year preceding forecast year-1 on November 1. To illustrate how the forecast years are defined, forecast year-1 is highlighted in red and forecast year-6 is highlighted in blue.

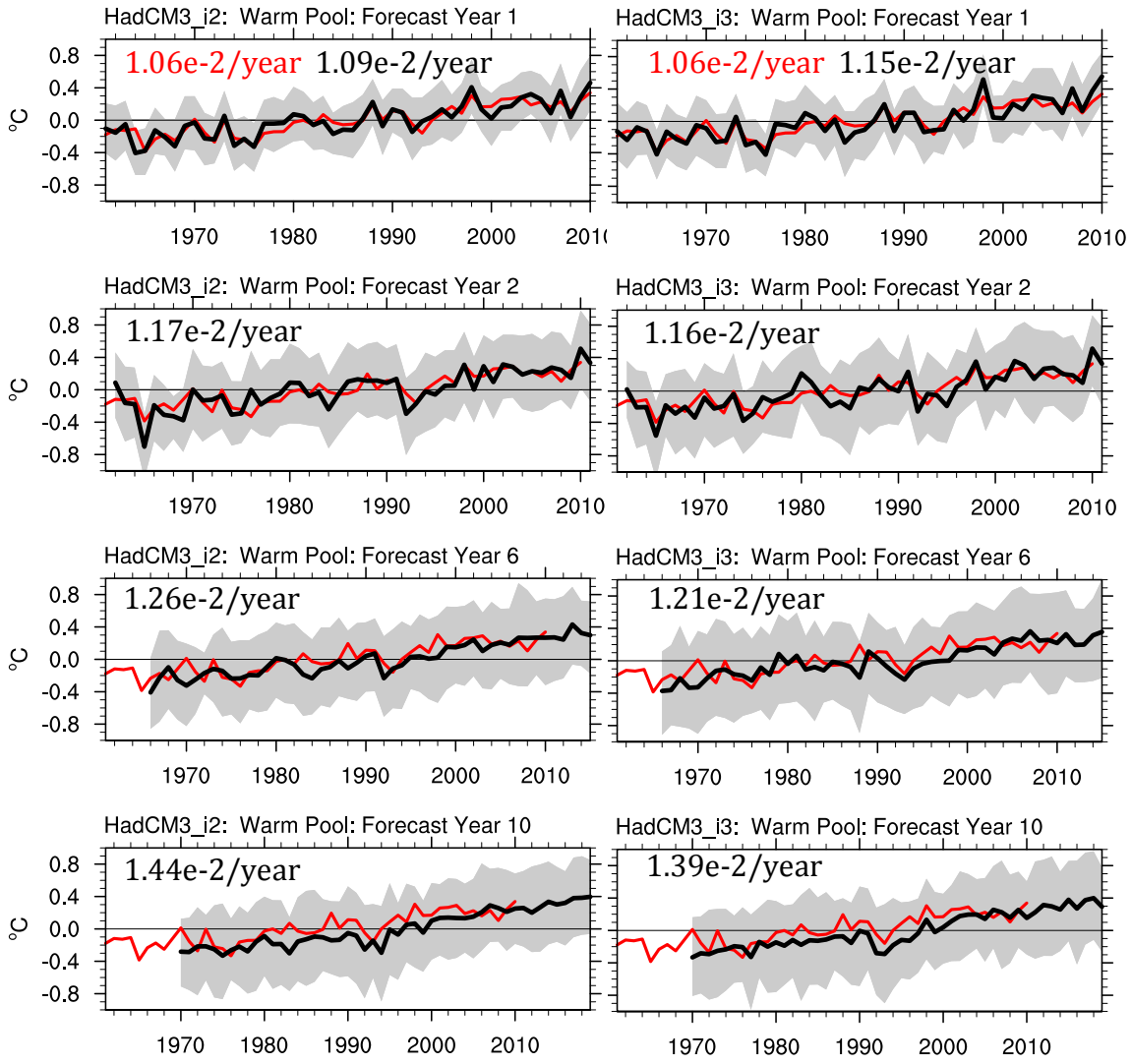

Figure 4: HadCM3 i2 (left) and HadCM3 i3 (right) Warm Pool index, in units of °C. Ensemble mean (thick black line) and ensemble spread (gray shading) for lead-time of 1, 2, 6, and 10 years. Observations shown with red line. Numbers show linear trend of the ensemble mean.

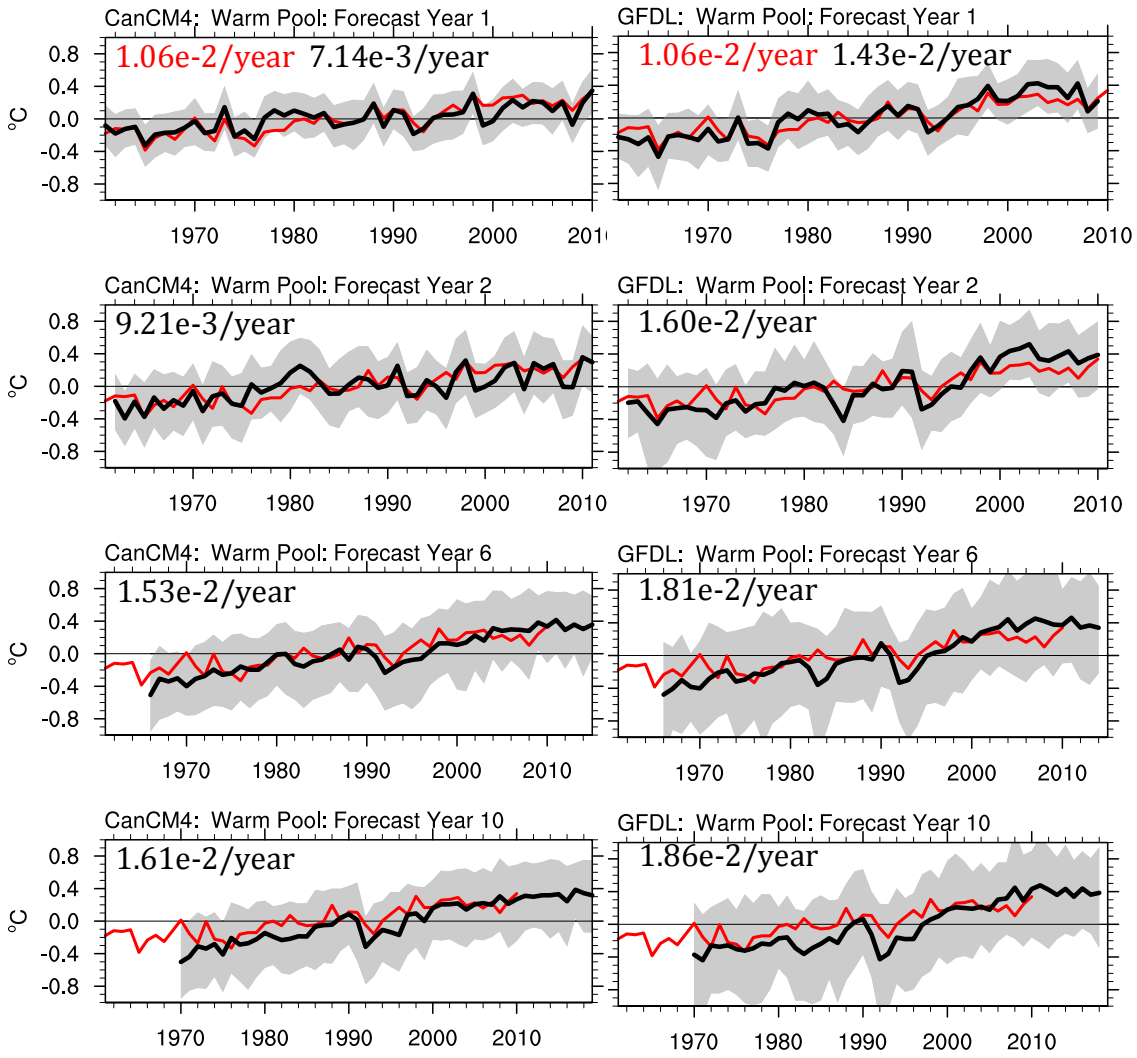

Figure 4 continued: CanCM4 (left) and GFDL (right) Warm Pool index.

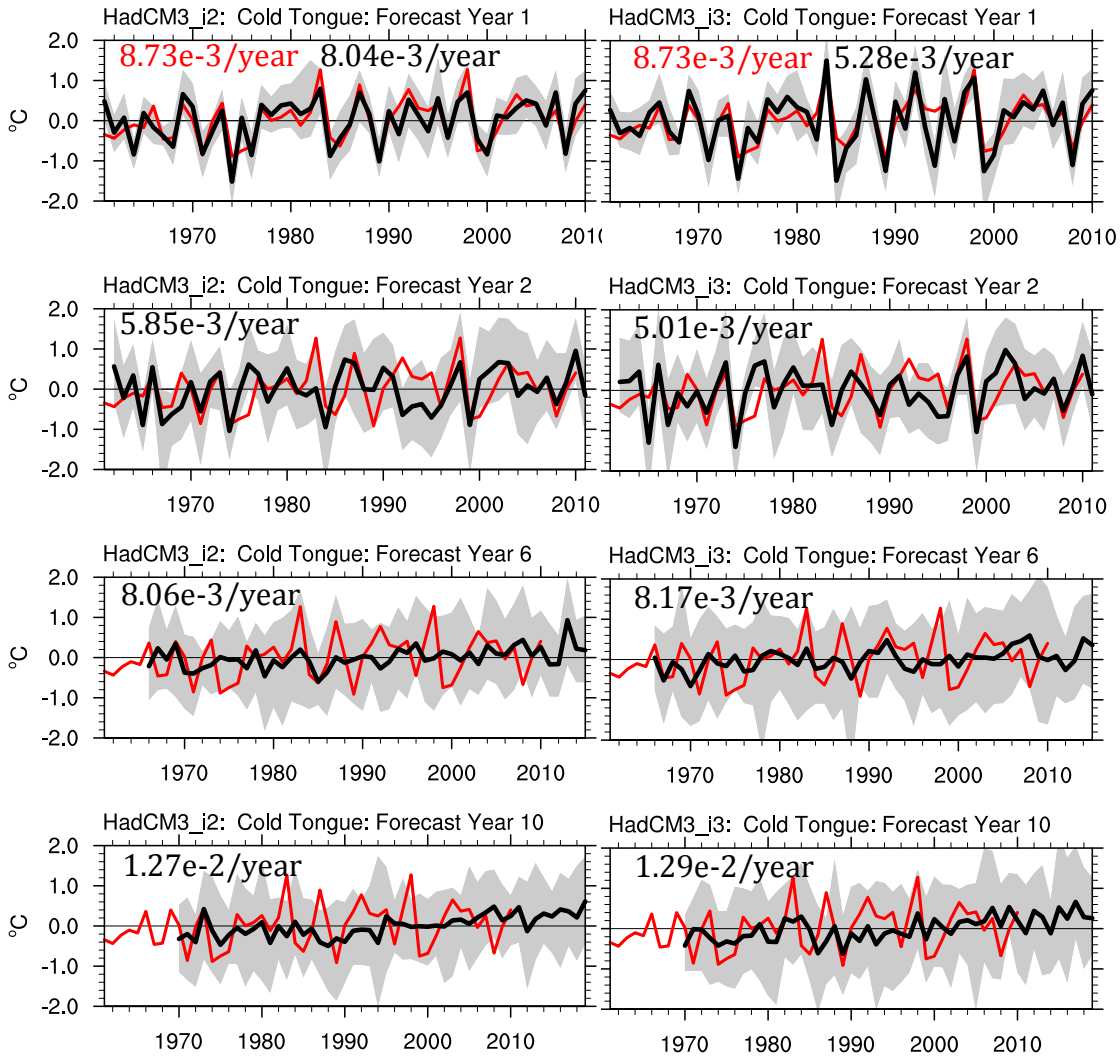

Figure 5: HadCM3 i2 (left) and HadCM3 i3 (right) Cold Tongue index, in units of °C. Ensemble mean (thick black line) and ensemble spread (gray shading) for lead-time of 1, 2, 6, and 10 years. Observations shown with red line. Numbers show linear trend of the ensemble mean.

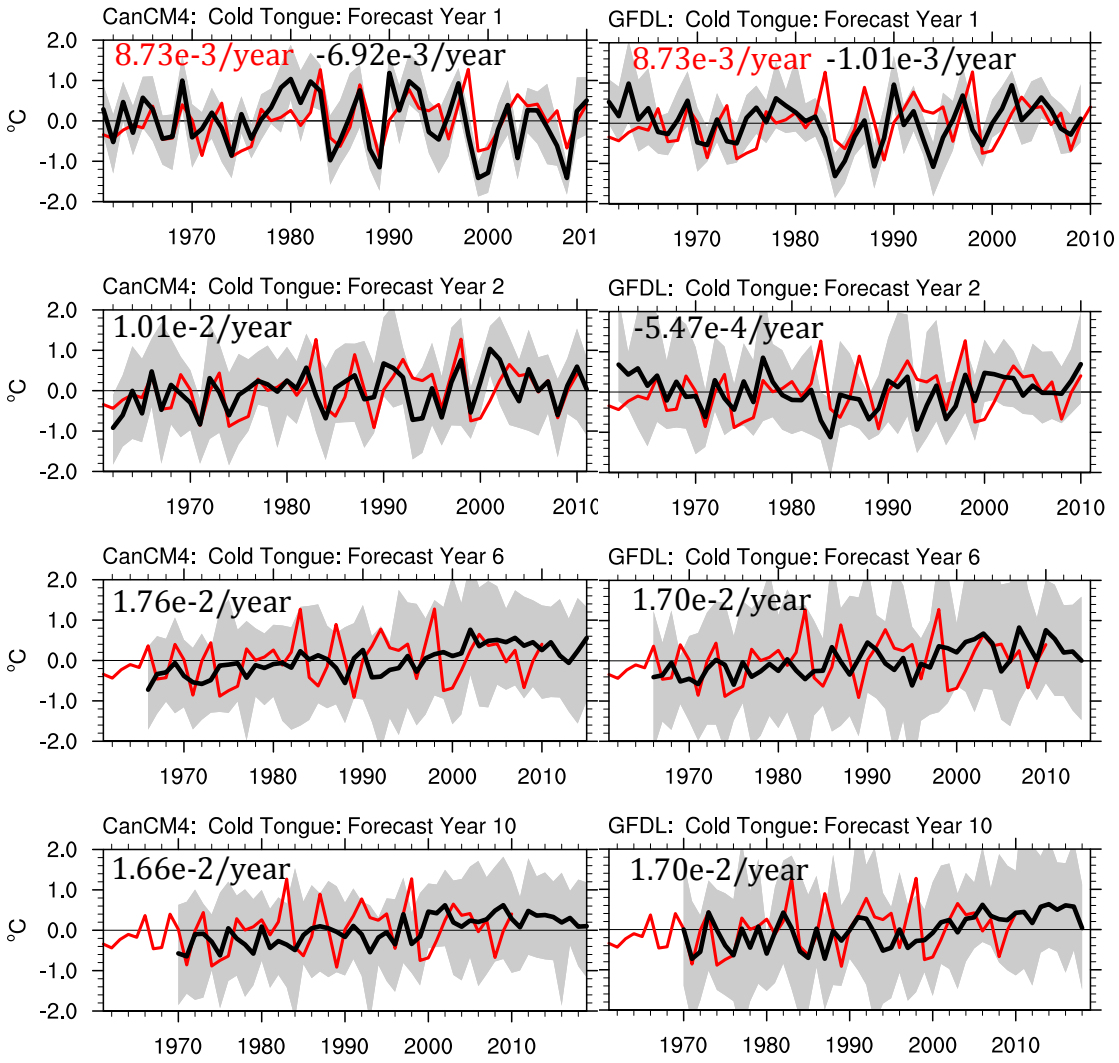

Figure 5 continued: CanCM4 (left) and GFDL (right) Cold Tongue index.

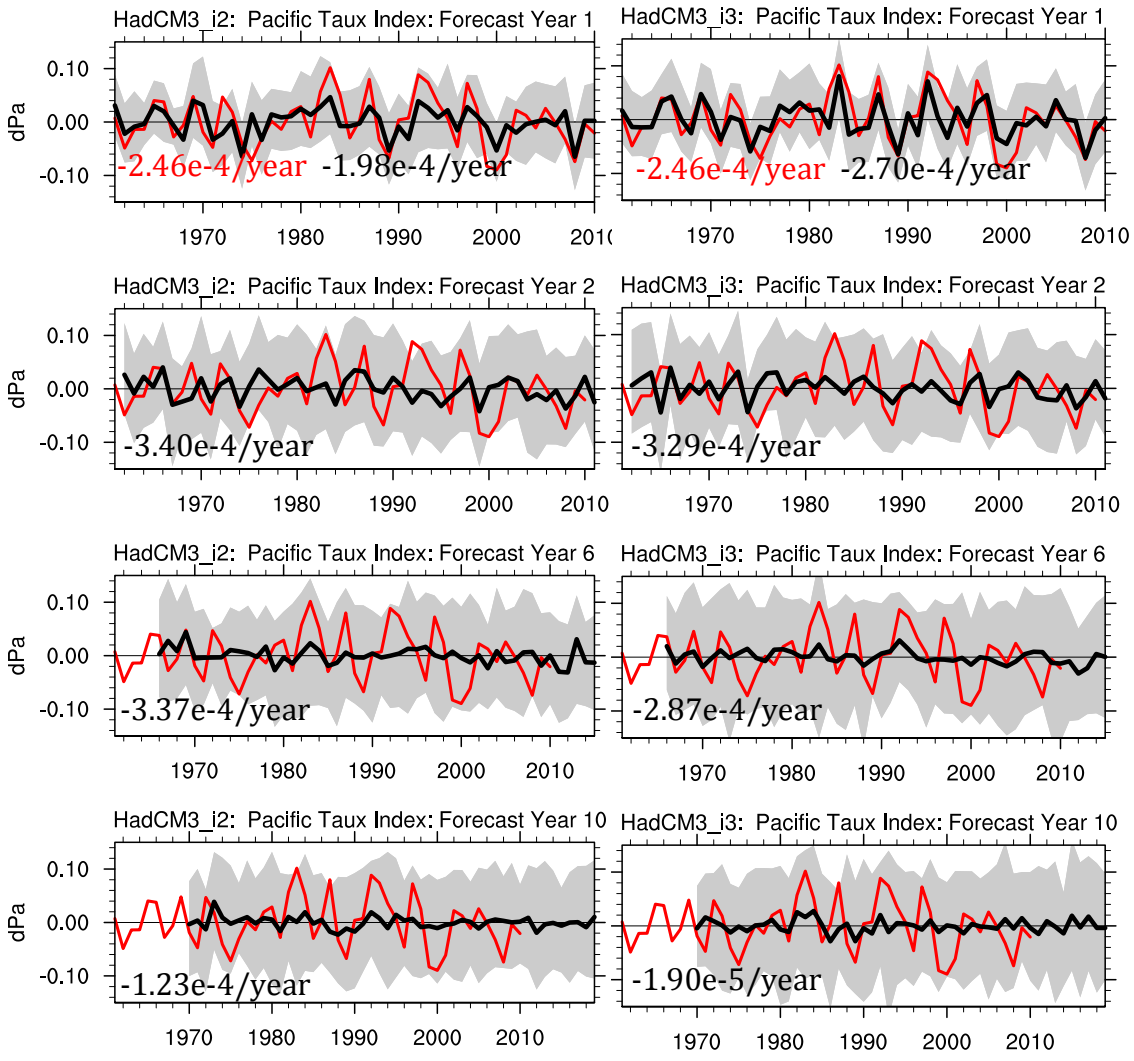

Figure 6: HadCM3 i2 and HadCM3 i3 Pacific τ_x index, in units of dPa. Ensemble mean (thick black line) and ensemble spread (gray shading) for lead-time of 1, 2, 6, and 10 years. Observations shown with red line. Numbers show linear trend of the ensemble mean.

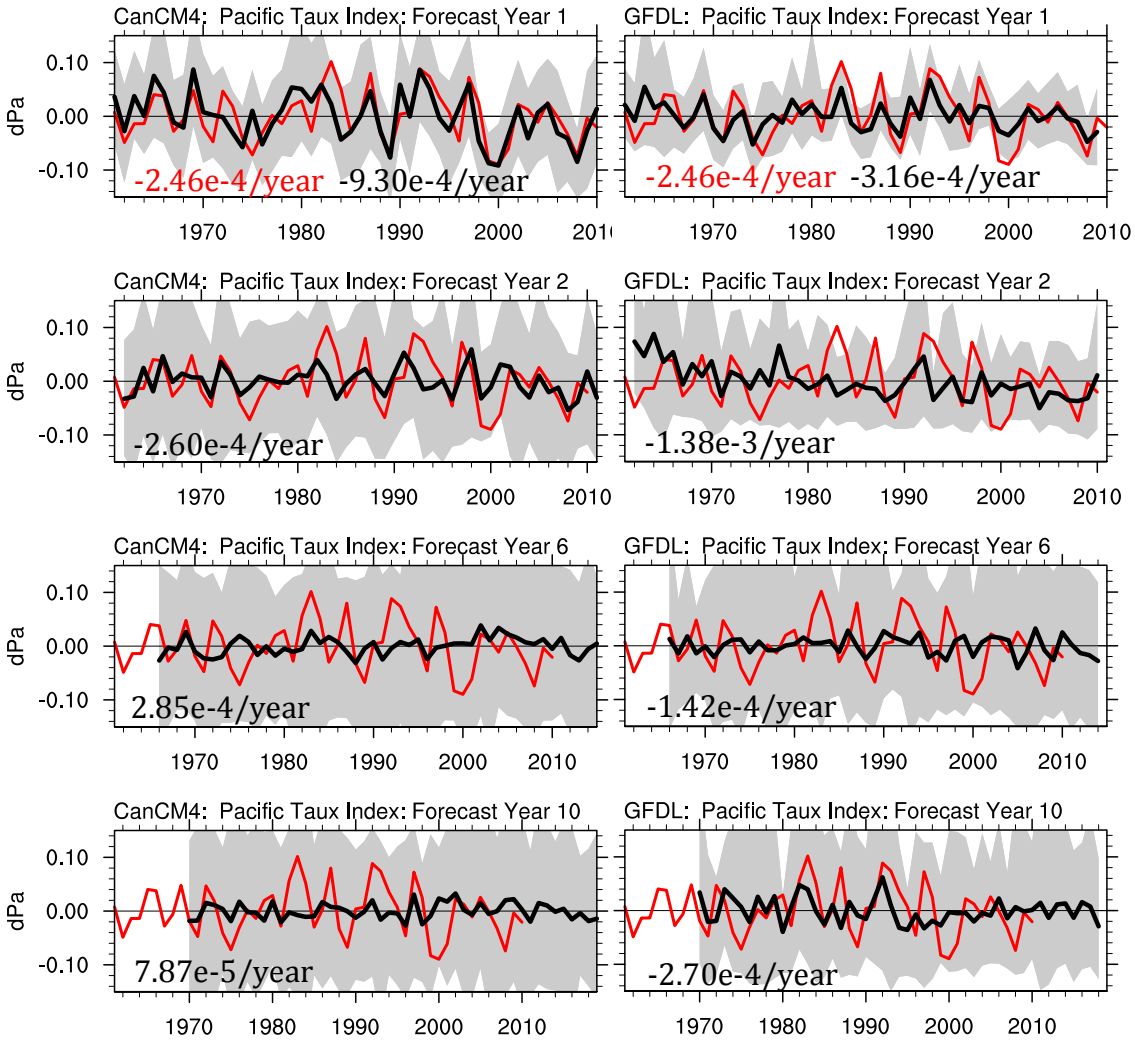

Figure 6 continued: CanCM4 and GFDL Pacific τ_x index.

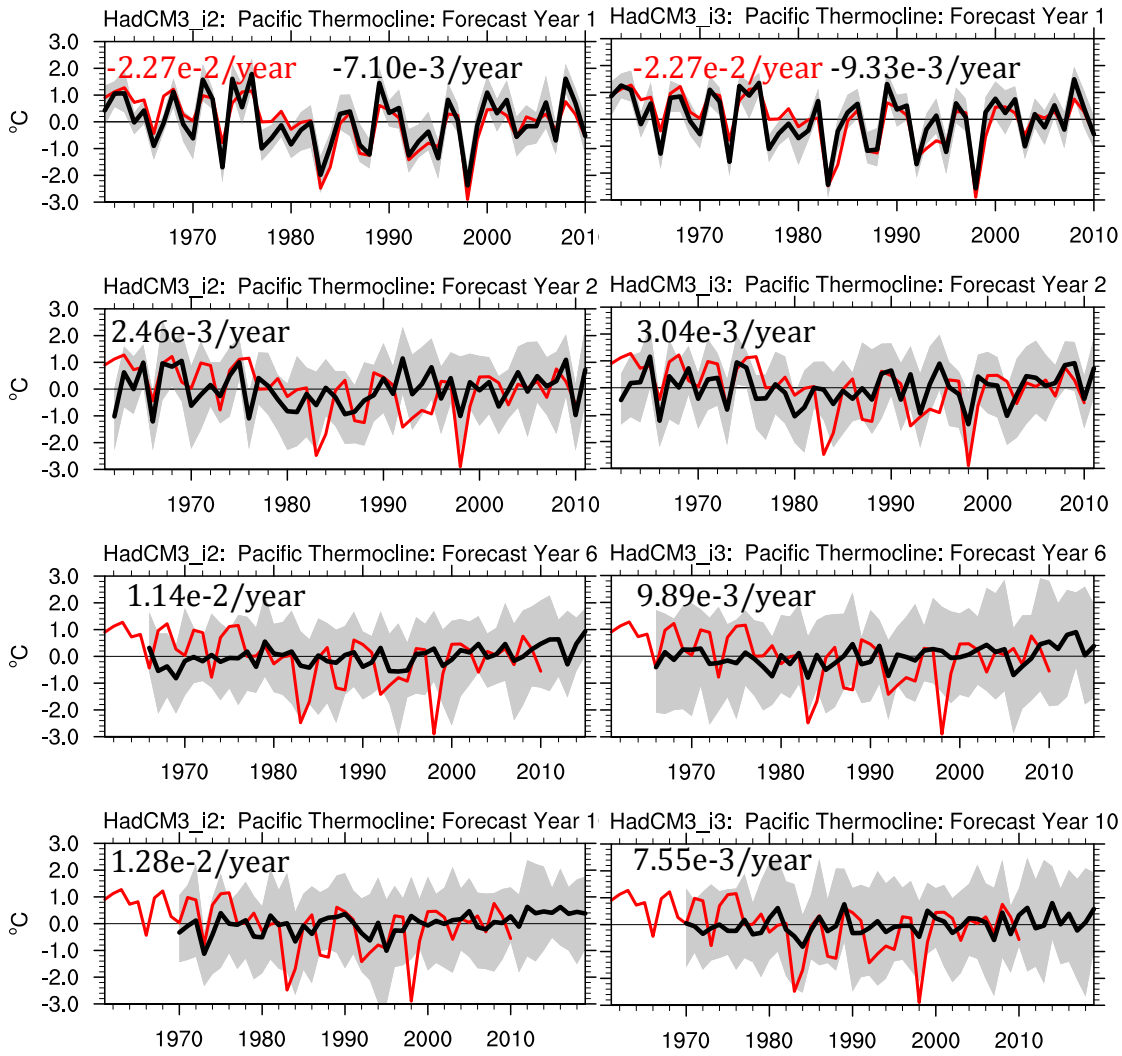

Figure 7: HadCM3 i2 (left) and HadCM3 i3 (right) Pacific Ocean thermocline index, in units of °C. Ensemble mean (thick black line) and ensemble spread (gray shading) for lead-time of 1, 2, 6, and 10 years. Observations shown with red line. Numbers show linear trend of the ensemble mean.

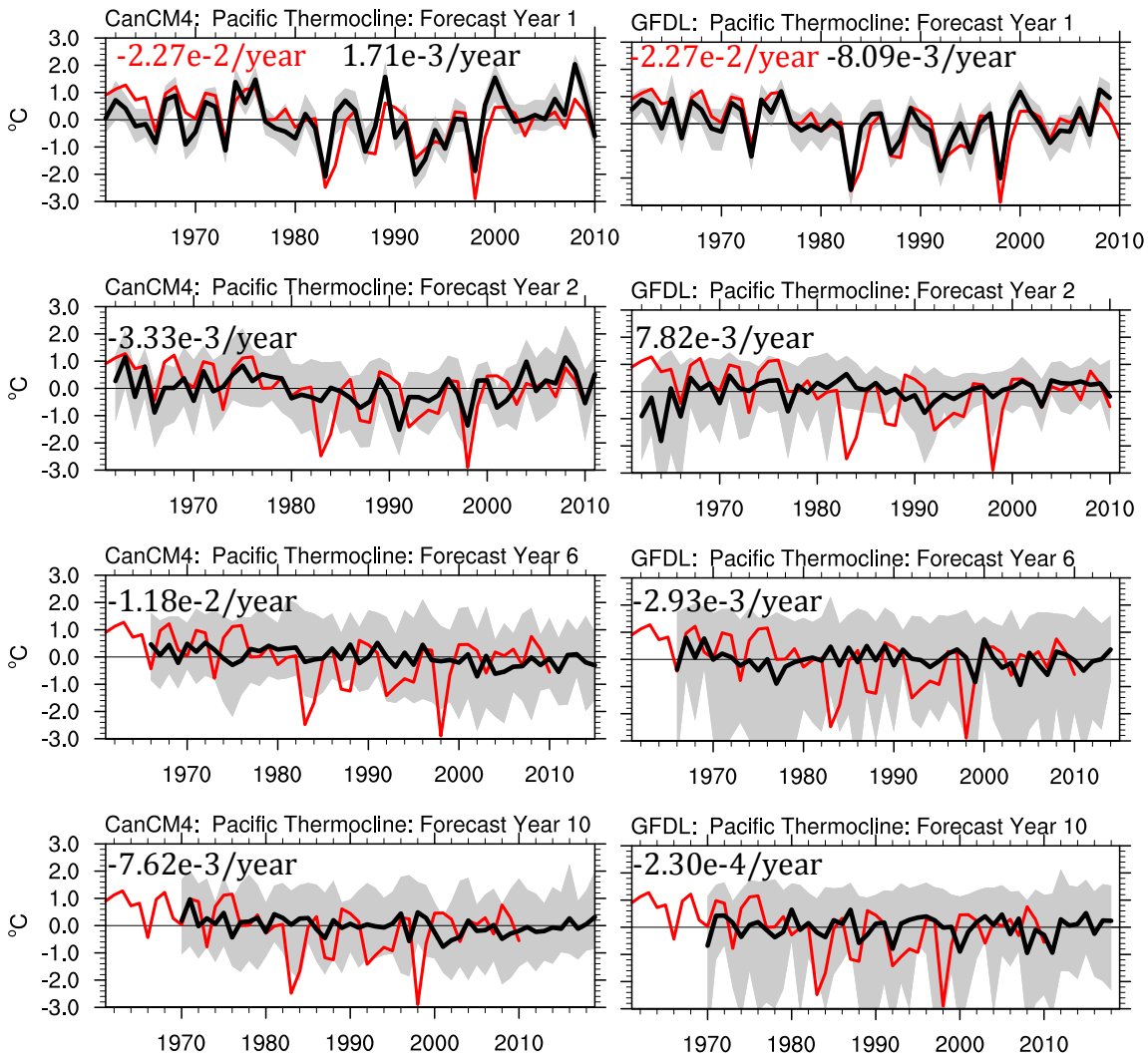

Figure 7 continued: CanCM4 (left) and GFDL (right) Pacific Ocean thermocline index.

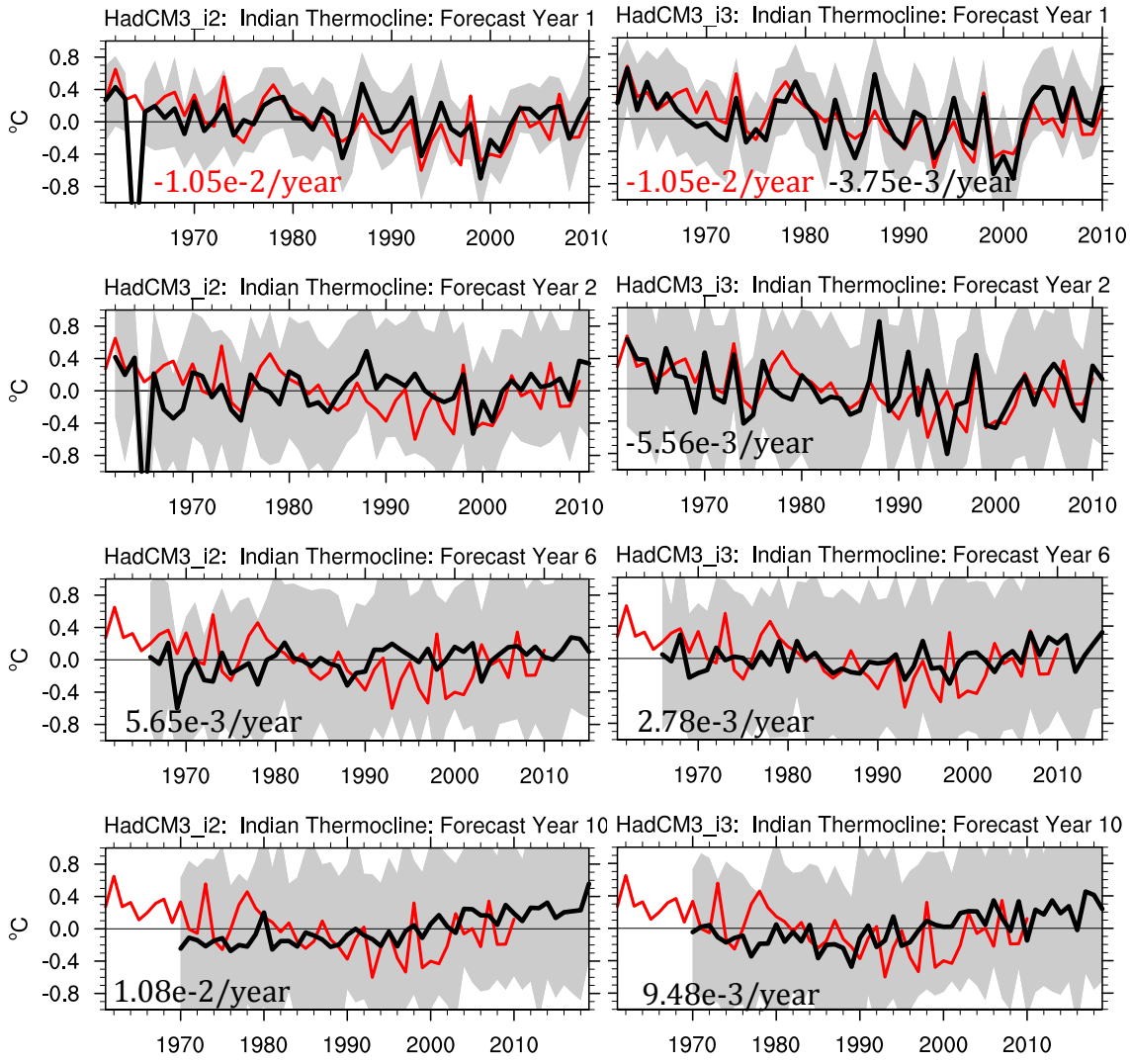

Figure 8: HadCM3 i2 (left) and HadCM3 i3 (right) Indian Ocean thermocline index, in units of °C. Ensemble mean (thick black line) and ensemble spread (gray shading) for lead-time of 1, 2, 6, and 10 years. Observations shown with red line. Numbers show linear trend of the ensemble mean.

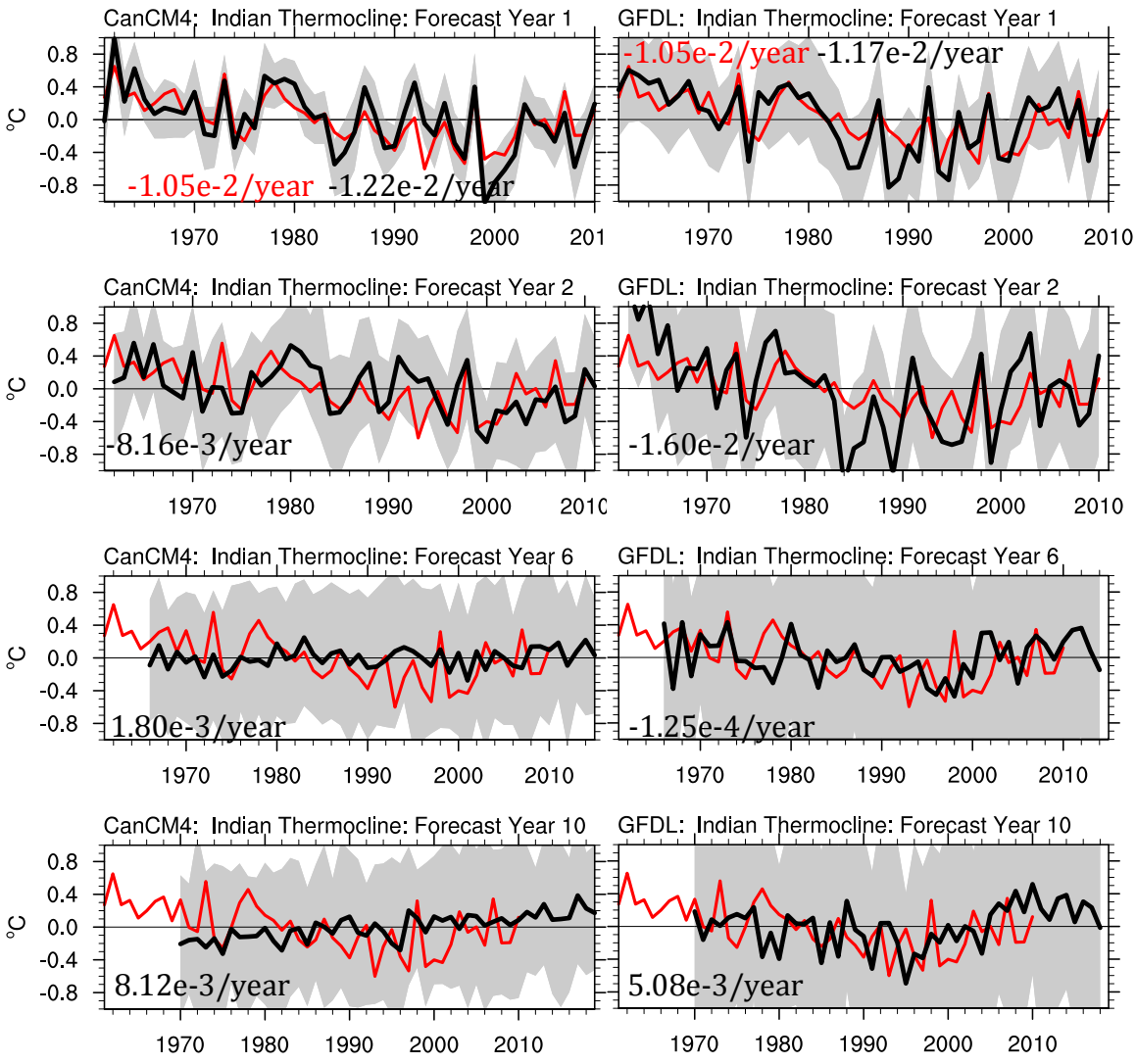

Figure 8 continued: CanCM4 (left) and GFDL (right) Indian Ocean thermocline index.

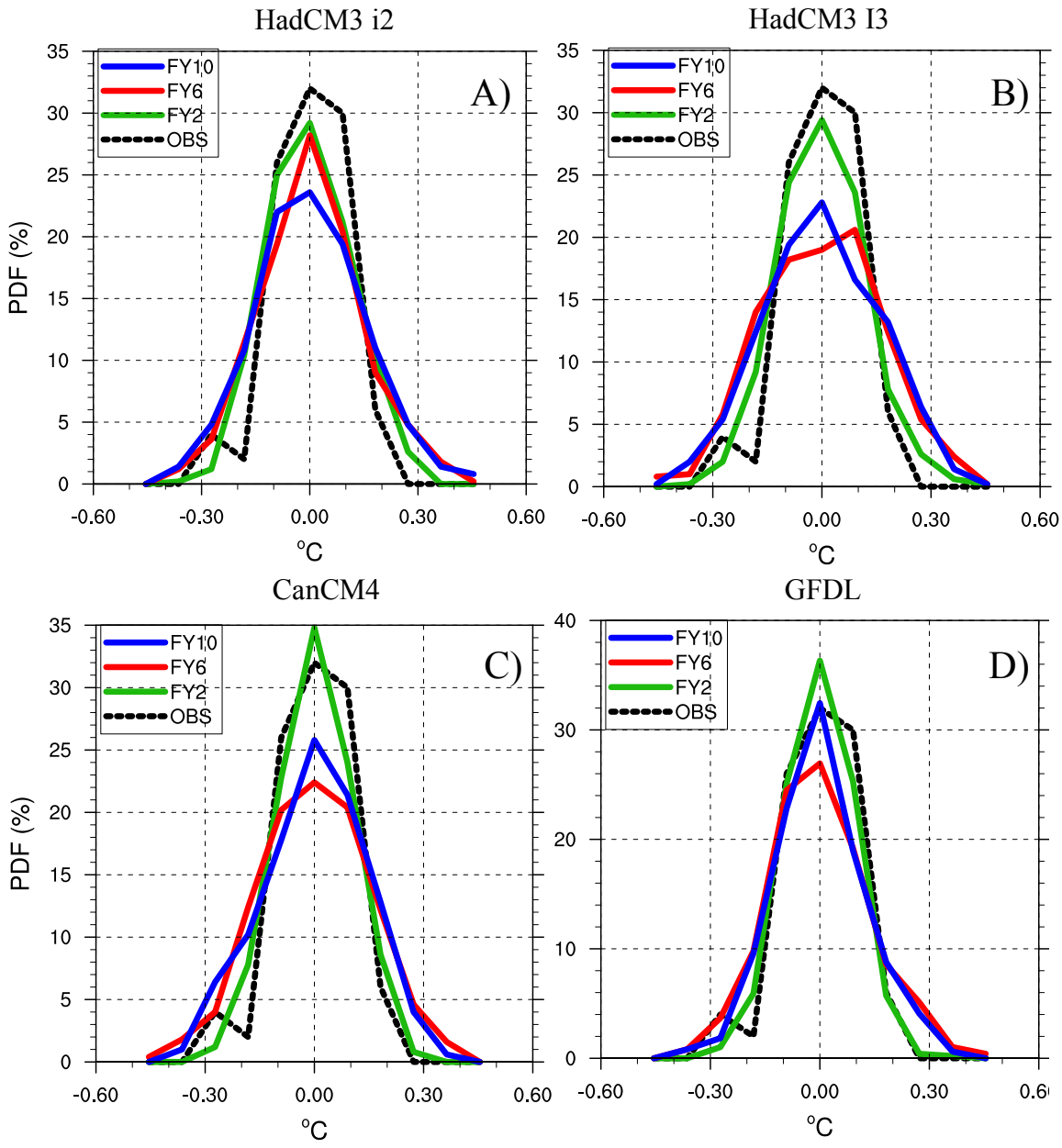

Figure 9: PDFs of detrended annual mean Warm Pool indices for lead-times of 2 years (green), 6 years (red), and 10 years (blue), in units of °C. PDFs of detrended observed annual means shown with black dashed line. A) HadCM3 i2. B) HadCM3 i3. C) CanCM4. D) GFDL.

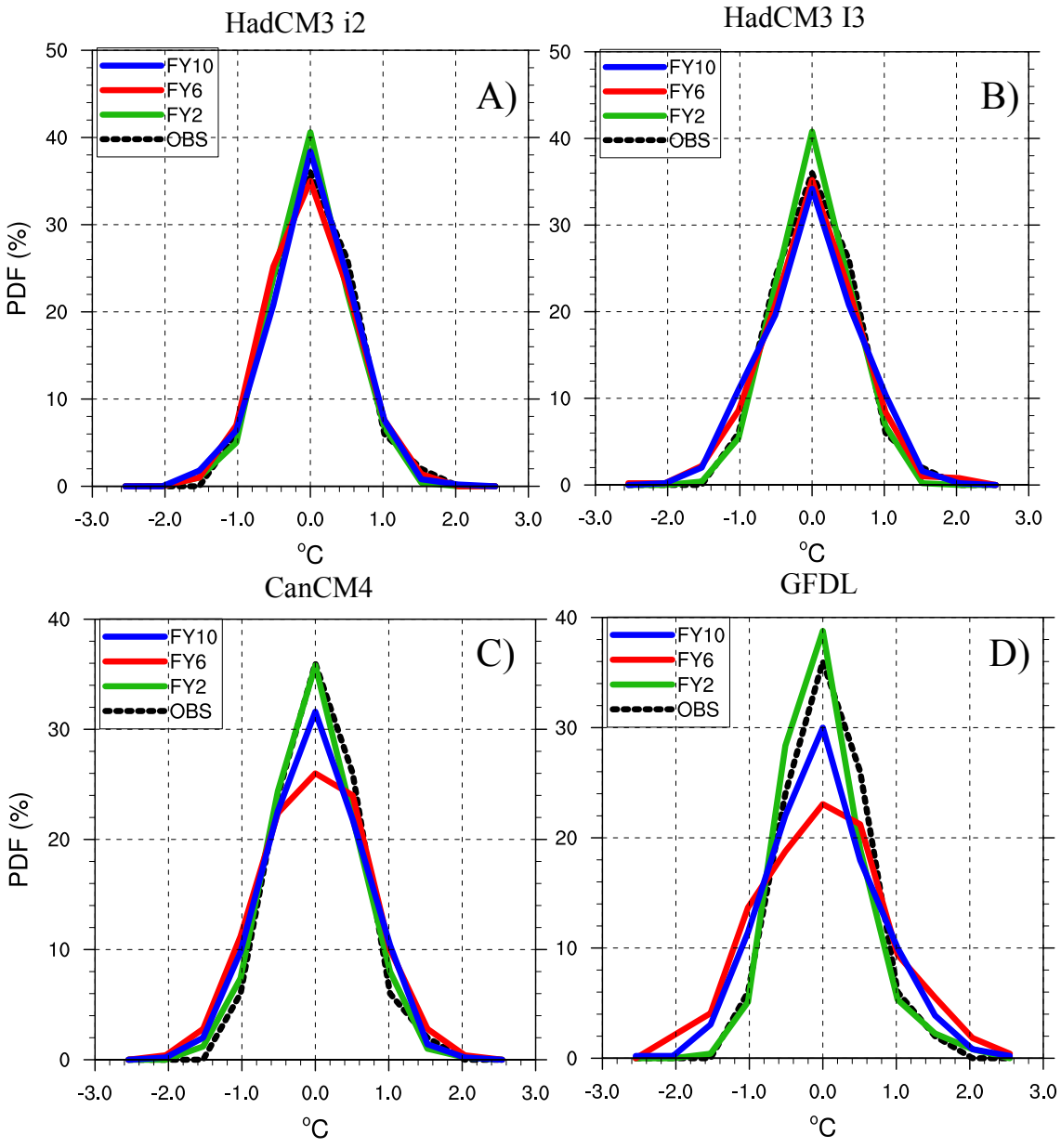

Figure 10: PDFs of detrended annual mean Cold Tongue indices for lead-times of 2 years (green), 6 years (red), and 10 years (blue), in units of °C. PDFs of detrended observed annual means shown with black dashed line. A) HadCM3 i2. B) HadCM3 i3. C) CanCM4. D) GFDL.

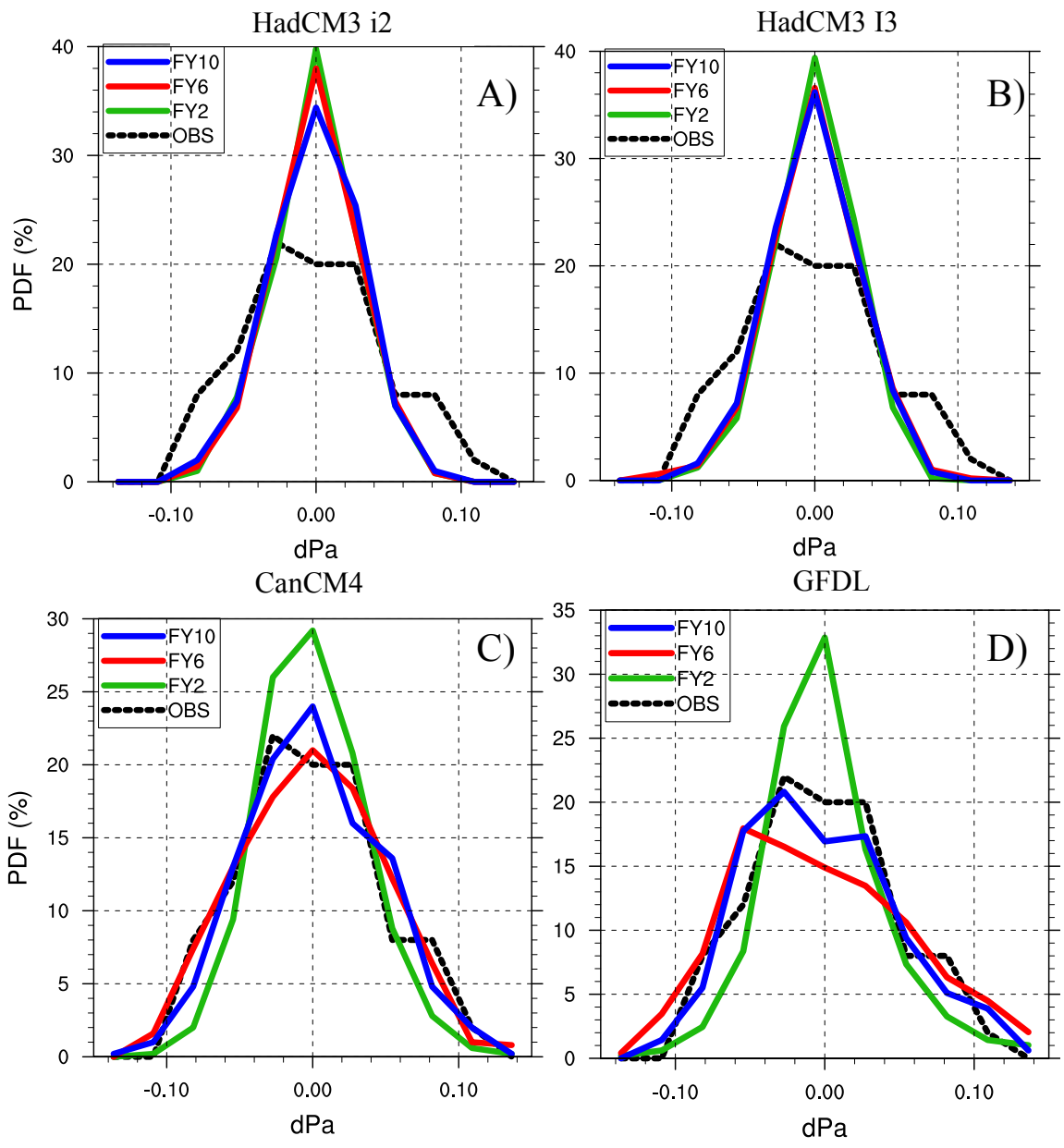

Figure 11: PDFs of detrended annual mean Pacific τ_x indices for lead-times of 2 years (green), 6 years (red), and 10 years (blue), in units of dPa. PDFs of detrended observed annual means shown with black dashed line. A) HadCM3 i2. B) HadCM3 i3. C) CanCM4. D) GFDL.

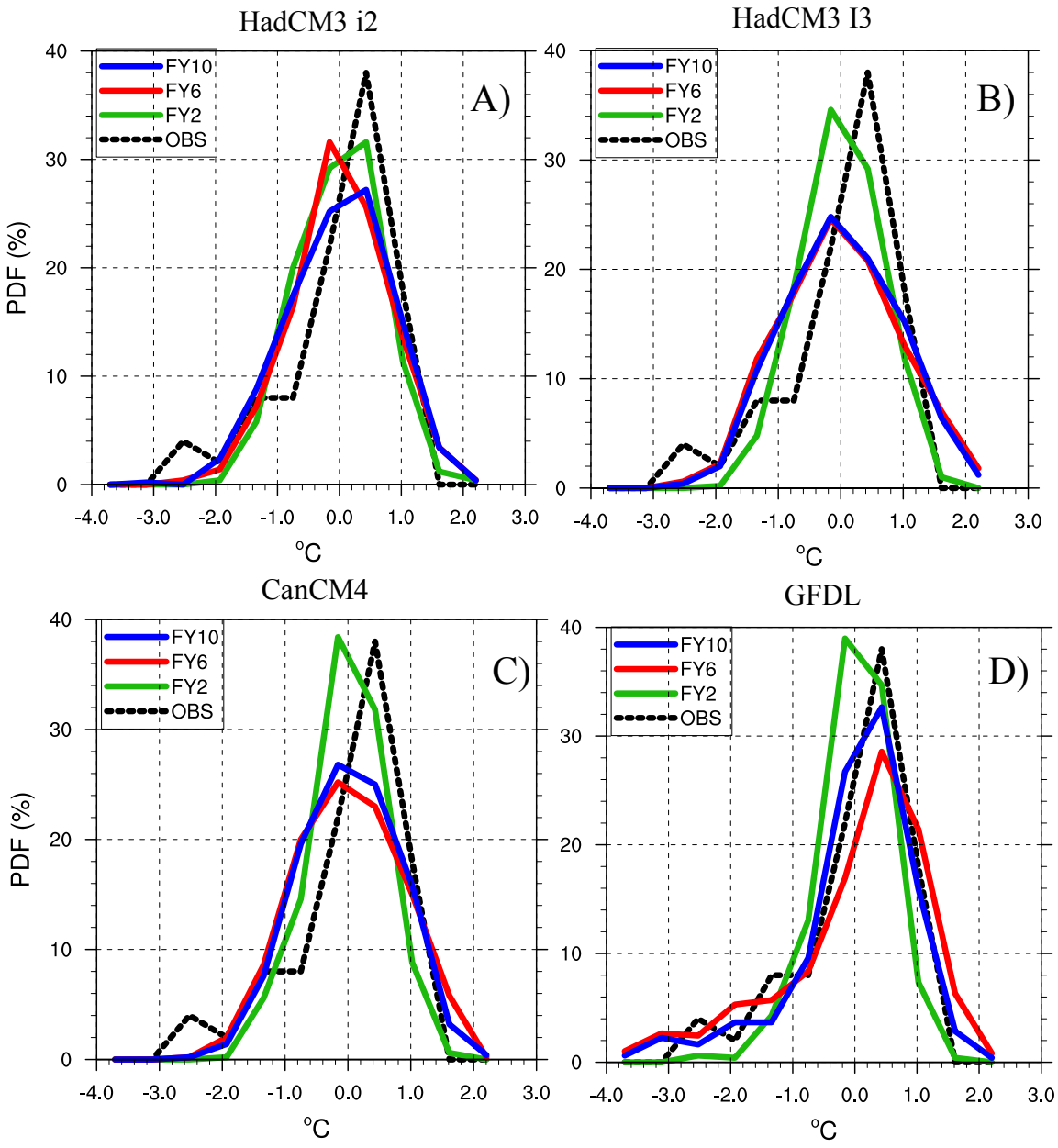

Figure 12: PDFs of detrended annual mean Pacific Ocean thermocline indices for lead-times of 2 years (green), 6 years (red), and 10 years (blue), in units of °C. PDFs of detrended observed annual means shown with black dashed line. A) HadCM3 i2. B) HadCM3 i3. C) CanCM4. D) GFDL.

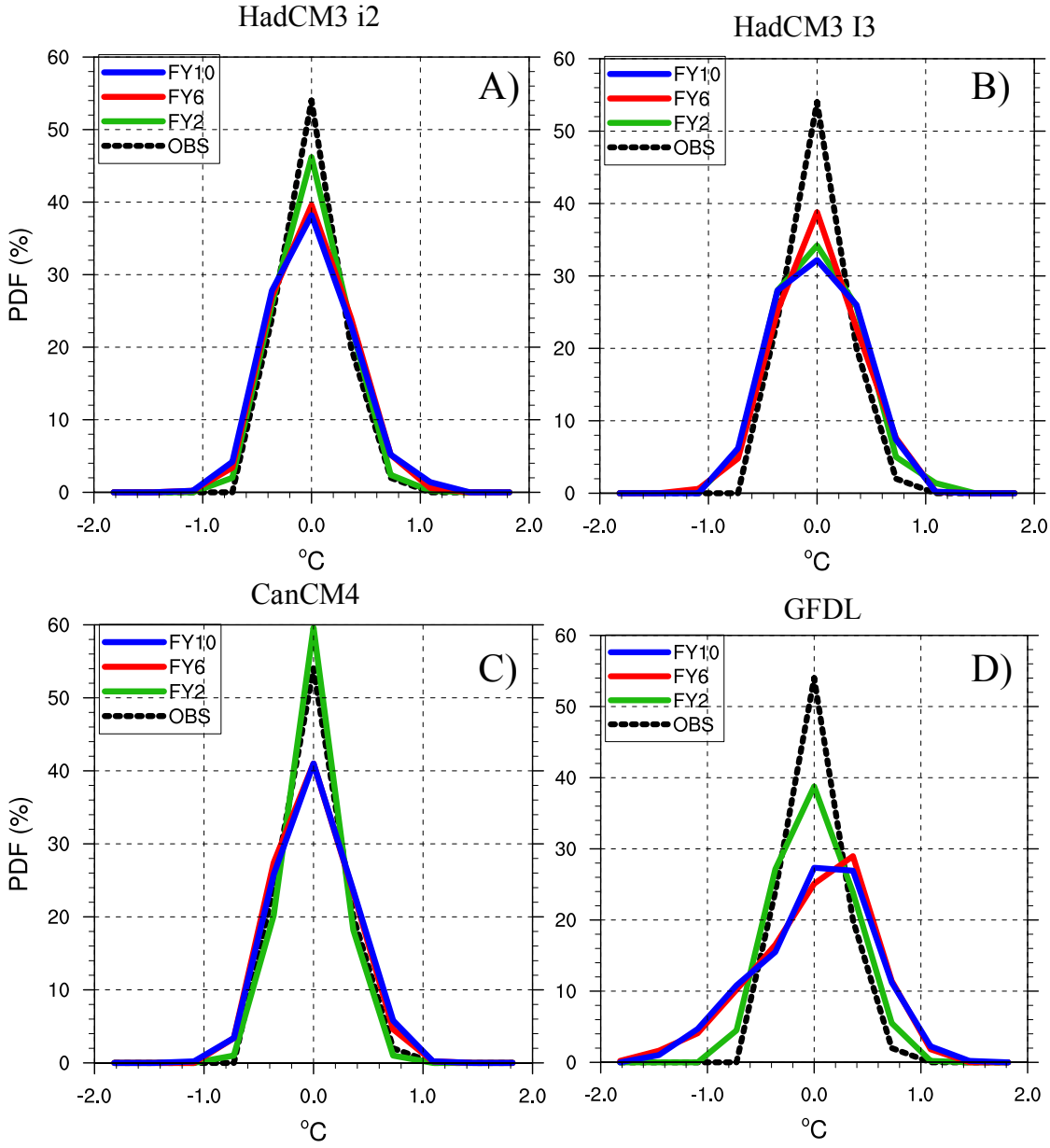

Figure 13: PDFs of detrended annual mean Indian Ocean thermocline indices for lead-times of 2 years (green), 6 years (red), and 10 years (blue), in units of °C. PDFs of detrended observed annual means shown with black dashed line. A) HadCM3 i2. B) HadCM3 i3. C) CanCM4. D) GFDL.